NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3543

Corresponding Measures:

De.2. Measure Title: Patient-Centered Contraceptive Counseling (PCCC) measure

Co.1.1. Measure Steward: University of California, San Francisco

De.3. Brief Description of Measure: The PCCC is a four-item patient-reported outcome performance measure (PRO-PM) designed to assess the patient-centeredness of contraceptive counseling at the individual clinician/provider and facility levels of analysis. Patient-centeredness is an important component in all areas of health care, and is uniquely critical in the personal and intimate process of contraceptive decision-making. The PCCC is intended to provide health care organizations with a tool to measure the quality of interpersonal communication between clinician/provider and patient – a core aspect of patient-centeredness – in the context of contraceptive care specifically.

The PCCC is specifically designed to capture three key domains of contraceptive care quality, as described as high priorities by patients themselves in previous qualitative research conducted by our team [1]. These domains include interpersonal connection between health care provider and patient, support in the contraceptive decision-making process, and adequate information to make such a decision. The four-item PCCC captures the three domains of quality contraceptive quality and retains validity and reliability of the original 11-item scale. Patients are asked to rate how well their individual health care provider did at each of the following, with each item presented on a 5-point Likert scale with responses ranging from 1 ("Poor") to 5 ("Excellent"):

- Respecting me as a person
- Letting me say what matters to me about my birth control
- Taking my preferences about my birth control seriously
- Giving me enough information to make the best decision about my birth control method

The target population for the PCCC is patients age 15-45, who were assigned female at birth, and who have received contraceptive counseling as part of their recent visit. The PCCC is visit-specific, and is given to patients who have been identified as having received contraceptive counseling during their visit.

An individual provider's score is determined by the proportion of patients who gave the highest rating for all four question on the survey. Likewise, a facility's score is calculated as the percentage of facility patients who gave the highest rating for all four questions.

References

[1] Dehlendorf C, Kimport K, Levy K, Steinauer J. A qualitative analysis of approaches to contraceptive counseling. Perspectives on Sexual and Reproductive Health. 2014;46(4):233-240.

1b.1. Developer Rationale: The PCCC is designed to give health care organizations, facilities, and providers the opportunity to measure the quality of their patients' experience of contraceptive counseling and implement quality improvement strategies to improve patient experience as needed. While PCCC results are intended to have stand-alone value to organizations, we also intend for the PCCC to serve as a balancing measure for currently endorsed measures of contraceptive provision (NQF measures #2903 and #2904), of which the Office of Population Affairs (OPA) is the steward. Below, we describe the rationales for a measure of patient experience of contraceptive counseling in its own right, and the rationale for this measure's use alongside contraceptive provision measures.

Rationale for a measure of patient experience of contraceptive counseling

As described in our Evidence Attachment, patient experience of contraceptive counseling is an important outcome in and of itself, in that it is highly valued by patients [1] and measures a core aspect of quality care – patient-centeredness – as defined by the Institute of Medicine in its report Crossing the Quality Chasm [2]. This is consistent with the National Quality Forum's consideration of patient experience as one of the domains of Patient Reported Outcomes, as described in the measure evaluation criteria guidance document [3]. In order to capture this outcome with a PRO-PM, we engaged in a process of measure development that was continually informed by the input of patients on their needs and preferences for this type of care. Therefore, the resulting PCCC measure allows for identification of whether patients are experiencing high quality care as they themselves define it. In addition, patient-centeredness of contraceptive counseling as measured by the PCCC has been demonstrated to be associated with contraceptive continuation at six months [4], indicating a relationship between patient experience of counseling and the ability of patients to achieve their own reproductive goals, including pregnancy prevention. Patient experience has also been linked to improved engagement with care in various contexts [5,6]; in the context of contraceptive care, this means that patients who receive patient-centered care may feel more able to continue engaging with the reproductive health care system not only for contraception, but also if and when they become pregnant and/or give birth [7]. As such, positive patient experience of contraceptive counseling can support positive pregnancy and birth outcomes such as reduced maternal mortality.

Given the important implications of patient-centeredness of contraceptive counseling, both for patient experience and reproductive health outcomes, many health care organizations are understandably invested in gathering information on the experiences of their patients and improving those experiences as needed. The PCCC serves as a tool that organizations can use to understand the patient-centeredness of counseling and evaluate quality improvement interventions. CAHPS measures have been used to monitor the effectiveness of educational interventions for providers and staff to improve patient experience [8,9]. Similarly, the PCCC may be used to inform quality improvement activities, such as contraceptive counseling training or implementation of decision support tools to support contraceptive decision making, and to track their impact over time [10].

Rationale for use alongside contraceptive provision measures

The motivation behind the development of the PCCC originated during OPA's development of measures #2903 and #2904, which focus on most and moderately effective (MME) contraception and long-acting reversible contraceptive (LARC) methods. The OPA team and others involved in the measure development process foresaw that use of these important measures could have the unintended consequence of incentivizing provider pressure on patients to use more effective methods. During the NQF endorsement process, this concern was voiced by stakeholders, including the National Partnership for Women & Families (NPWF). The NPWF submitted a public comment that stated, "It is extremely important to keep in mind that reproductive coercion has a troubling history, and remains an ongoing reality for many, including low-income women, women of color, young women, immigrant women, LGBT people, and incarcerated women. We hope this measure will be paired with a woman-reported 'balancing measure' of experience of receiving contraceptive care. Such a measure can be expected to help identify and/or check inappropriate pressure from the health care system." Following endorsement of the contraceptive provision measures by the NQF, OPA acted

on this shared concern by funding a three-year cooperative agreement with UCSF in order to develop a PRO-PM as a 'balancing measure' to support proper use of the provision measures, and to enable health facilities and systems to measure patient experience in its own right. This initial funding supported our team at UCSF to work to reduce the IQFP to become the PCCC, as described in our Evidence Attachment. Further private foundation funding has supported the PCCC's validity and reliability testing in health care settings across the country.

Our team at UCSF intends to conduct further work to optimize use of the provision measures and the PCCC together. We hypothesize that the PCCC will serve as balancing measure for the provision measures, so that organizations can observe any fluctuations in PCCC scores that occur in association with changes in provision scores, and ensure that any increased provision does not come at the cost of patient experience, or ideally would in fact be associated with improved patient experience. As such, use of these two types of measures together can result in a more robust picture of contraceptive care quality, and assist health care organizations to achieve both aspects of quality in contraceptive care: providing access to a range of contraceptive methods and providing patient-centered counseling free of coercion.

References

[1] Dehlendorf C, Levy K, Kelley A, Grumbach K, Steinauer J. Women´s preferences for contraceptive counseling and decision making. Contraception. 2013;88(2):250-256.

[2] Wolfe A. Institute of Medicine Report: crossing the quality chasm: a new health care system for the 21st century. Policy, Politics, & Nursing Practice. 2001;2(3):233-235.

[3] National Quality Forum. Measure evaluation criteria and guidance for evaluating measures for endorsement. 2018. http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2018_Measure_Evaluation_Criteria_and_Guidance.aspx. Accessed 31 Jul 2019.

[4] Dehlendorf C, Henderson JT, Vittinghoff E, et al. Association of the quality of interpersonal care during family planning counseling with contraceptive use. American Journal of Obstetrics and Gynecology. 2016;215(1):78. e71-78. e79.

[5] Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. Medical Care Research and Review. 2014;71(5):522-554.

[6] Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open. 2013;3(1):e001570.

[7] Gomez AM, Wapman M. Under (implicit) pressure: young Black and Latina women´s perceptions of contraceptive care. Contraception. 2017;96(4):221-226.

[8] Horton DJ, Yarbrough PM, Wanner N, Murphy RD, Kukhareva PV, Kawamoto K. Improving physician communication with patients as measured by HCAHPS using a standardized communication model. American Journal of Medical Quality. 2017;32(6):617-624.

[9] Briggs K, Sharma L, Chandrasekaran A, Douglas C, Aroh D, Finefrock D. The effect of a hybrid training program. Nursing Management. 2018;49(2):51-53.

[10] Dehlendorf C, Fitzpatrick J, Fox E, et al. Cluster randomized trial of a patient-centered contraceptive decision support tool, My Birth Control. American Journal of Obstetrics and Gynecology. 2019;220(6):565. e561-565. e512.

S.4. Numerator Statement: The PCCC is a visit-specific measure of patient-centeredness in contraceptive counseling. It specifically measures how many patients report a top-box (i.e., the highest possible) score of patient experience in their contraceptive counseling interaction with a health care provider during their recent visit.

S.6. Denominator Statement: The target population for the PCCC is patients age 15-45, who were assigned female at birth, who are not currently pregnant, and who received contraceptive counseling as part of their recent visit.

S.8. Denominator Exclusions: Pregnant patients are excluded from the denominator, based on two reasons. First, contraceptive counseling in the context of pregnancy is distinct from that provided to non-pregnant individuals. Specifically, perinatal contraceptive counseling often includes multiple conversations touches over the course of prenatal care and immediate postpartum care. This is appropriate as women, when pregnant, are not immediately at risk of an undesired pregnancy, and therefore there is less time sensitivity to this counseling, and is also consistent with women's preferences for this care [1]. Given this difference in structure of counseling for pregnant women, the use of a visit-specific measure for contraceptive counseling is not appropriate.

Second, given distinct issues related to post-partum contraceptive use, including increased risk of blood clots, effect on lactation, and the health impact of birth spacing, counseling pregnant women about future contraceptive use has components distinct from that of non-pregnant women. For these conceptual reasons, the PCCC was designed for use with non-pregnant patients and has not been extensively tested with pregnant patients to determine whether it accurately captures their needs and desires for counseling.

References

[1] Yee LM, Farner KC, King E, Simon MA. What do women want? Experiences of low-income women with postpartum contraception and contraceptive counseling. Journal of Pregnancy and Child Health. 2015;2(5).

De.1. Measure Type: Outcome: PRO-PM

S.17. Data Source: Instrument-Based Data

S.20. Level of Analysis: Clinician : Individual, Facility

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If these data are not available, data should demonstrate wide variation in performance and be from a robust number of providers with results that are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- The developer demonstrates a connection between the outcome (patient has a positive experience and receives patient-centered care) and both a healthcare structure (appropriate range of methods and appropriate counseling are available) and a process (counseling is provided; appropriate contraceptive is prescribed if needed).
- The developer states that "more generally, evidence supports the proposition that health care system interventions, including training and counseling interventions targeted at patient-provider communication can produce change in patient experience outcomes."
- This measure focuses on the interaction between providers and patients, which is a process that can be improved. The developer notes that "Processes or interventions influencing this interaction therefore have the potential to influence scores on this measure." It notes that tools, such as the evidence-based shared decisionmaking support tool, *My Birth Control*, can be used to help providers ensure patient choices are

centered. The developer notes that "Using data from a randomized controlled trial of 749 women in the San Francisco Bay Area, we found that use of *My Birth Control* was associated with a top-box response to the PCCC (with 72.4% of intervention-arm participants giving a top-box response on the PCCC vs. 64.7% of control-arm participants, p=0.026 [unpublished data])."

- The developer also notes that higher scores on the PCCC are associated with contraceptive continuation and that "qualitative data suggests that patients who experience non-patient-centered care are less likely to return to seek out care for future reproductive health needs. This has the potential to negatively impact a range of outcomes, including pregnancy-related morbidity and mortality."
- To demonstrate that the target population values the measured outcome, the developer conducted 42 in-person interviews and a modified grounded theory of qualitative analysis. From this qualitative data, it concludes that patients find value in having positive experiences with contraceptive counseling. Specifically, the developer states that patients value "interpersonal connection between provider and patient, adequate information, and decision support", which this measure aims to capture.
- This measure was originally conceptualized to be a balancing measure for *Contraceptive Care – Most & Moderately Effective Methods* (NQF #2903) and *Contraceptive Care – Access to LARC* #2904, to ensure that providers are not incentivized to reduce quality of patient-centered care while they are increasing quality of LARC services offered. However, as a patient experience of care measure, it can also be used to monitor patient experiences to ensure patient-centered care is being provided.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *This measure is derived from patient report. Does the target population value the measured outcome and find it meaningful?*

Guidance from the Evidence Algorithm

Measure assesses a PRO (Box 1) → Relationship between measured PRO and healthcare action is demonstrated (Box 2) → Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The variability in scores reported by the developer suggests an opportunity for improvement in low performers. Developer also notes that its score distribution is wider than the CG-CAHPS communication composite score.
- Provider-level analysis (n=34 providers who provided counseling to 2,477 patients)
  - Mean performance score: 0.81
  - Standard deviation: 0.12
  - Range: 0.44-0.95
  - Percentiles
  - 25th: 0.79
  - 50th: 0.85

- 75th: 0.90

- Facility-level analysis (n=22 facilities that provided counseling to 3,478 patients)
  - Mean: 0.79
  - Standard deviation: 0.12
  - Range: 0.51-0.97
  - Percentiles
  - 25th: 0.70
  - 50th: 0.83
  - 75th: 0.88
  - The developer also notes that "research conducted by our team at UCSF examining quality of counseling via audio recording of patient visits found that providers inconsistently elicited or engaged with patient experiences and preferences during counseling."

Disparities

- The developer reports that, among patients who received contraceptive counseling, there were differences in the frequency with which different racial and ethnic groups gave the top box score to providers, and that this varied across clinics. People of color and Spanish speaking individuals gave the top box score less frequently. The developer plans to provide data stratified by race and ethnicity upon maintenance, if endorsed. The developer addresses its decision not to risk adjust in the testing section.
- The overall percentage of top-box scores:
  - 78.5% for Black patients
  - 74.4% for Asian patients
  - 86.0% for White patients, and
  - 80.4% for Hispanic/Latina patients compared to 84.1% for non-Hispanic/Latina patients
  - 83.9% for English speakers vs. 68.2% for Spanish speakers

Question for the Committee:

- Is there a gap in care that warrants a national performance measure?

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Importance to Measure and Report

Comments:

**This appears to be direct evidence of patient-centered contraceptive care as reported by the patient. The developers considered this to be an outcome measure, but I might argue that this is more of a process measure-e.g. capturing the process of contraceptive counseling, not the outcome of the counseling. This also appears to have been validated across various target populations.

**The materials demonstrate the evidence to support that desired outcomes are related to the process/structure and that patients value the measured process/structure.

**The developers carried out qualititative research to identify factors important to women when receiving contraceptive care and used results to develop their PRO-PM. They have continued to carry out other research collecting input from women and documenting positive effects of use of the measure. In their rationale statement,

the developers associate the measure with patient-centeredness, a core dimension of quality in the Crossing the Quality Chasm framework. Conversely, women's dislike of contraceptive care that is disrespectul, including coercion and failing to meet their information needs, is well-documented, as is the fact that such care disproportionately is experienced by women of color, low-income women, immigrants, LGBTQ community members, incarcerated women, women with disabilities, and other marginalized and oppressed groups.

**The evidence supports the need to develop a patient-reported outcome measure. Considering the history of contraceptive coercion, especially for minority and low-income populations along with those with SMI, having a PRO is an important balancing measure.

**I believe that the developers have made a good case that patients value a positive interaction for contraceptive counseling. However, the data that a patient perceived positive interaction lowers the rate of unwanted pregnancy seems weak. They do associate a positive interaction with continued use of a contraceptive, but that is truely a process measure rather than an outcome.

**Evidence provided by the measure developer directly relates and supports the outcomes being measured. The structure, process and outcomes of the assessment/survey informs about the patient experience. The in person interviews and analysis proved that patients find value in the measure outcomes, process and structure.

**Patient centered care, often performed through motivational interviewing, is widely recognized as a vehicle for maximizing patient engagement in behavior change and optimizing a valued choice by the patient

**Not aware on new studies. agree with measures

**There is strong evidence and is directly applicable. How will this survey be provided to patients? What is the data burden for providers?

1b. Performance Gap

Comments:

**Yes, they provided performance data from a study in the Bay Area, and demonstrated a gap in care and variability based on language and race/ethnicity.

**This is a new measure meant to balance other contraceptive measures.

**Whether standalone or as a balancing measures with #2903 and #2904, there is a longstanding gap, worse for marginalized and oppressed groups, between the respectful care that women deserve and the quality of care they often get. Quality concerns include coercion, lack of adequate information, insensitive treatment for an intimate aspect of health care. These concerns are longstanding, well documented and definitely continuing to the present. The developers document varied performance on the measure.

**It will fulfill a need to ensure that people are informed, have their questions answered, and are not pressured into contraceptive services.

**Yes

**The developer provided data that showed variations in scores at the provider and practice level. It suggests there are opportunities for improvement with the lower performing providers.

**There was clearly a demonstrated gap in performance as demonstrated by the stated ranges. In looking at the developer's graphs, there was clear clumping of performance at the high end. In looking at the graphs, it appeared that the data could be used in a more bimodal fashion to intervene with low performers and encourage higher performers

**Yes gap was provided

**There is a performance gap related to race and ethnicity. Will the committee recommend formal bias training for providers?

1b. Disparities

Comments:

**Yes, data by race/ethnicity and language (eng/span) demonstrated disparities in care

**Evidence was provided that shows disparities in patient experience with contraceptive counseling between white and black women, and between Latina and non-Latina women.

**The developers report that average top-box score varies by race/ethnicity and language. Lower scores are in an expected direction considering a broad range of disparities, e.g., lower for Black than white women and for Spanish versus English speakers.

**The data was respectful of diverse populations; especially those who are at greatest risk.

**Good demonstration that interaction ratings vary by race/ethnicity

**Yes, data by population subgroups was provided and showed variations in top scores by ethnicity/race.

**The disparity data was interesting but would require different followup data or interview or observation to undertake performance improvement

**Yes

**Yes, there is discrepencies between Black/Asians vs Whites and English speaking pts vs non-English speaking patients.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: NQF Scientific Methods Panel Subgroup

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed and discussed by the Scientific Methods Panel (SMP). A summary of the SMP discussion is provided below.

Scientific Methods Panel Votes: Measure passes

- Reliability: H-5; M-1; L-0; I-0
- Validity: H-5; M-1; L-0; I-0

Reliability

- Reliability testing was performed at the data element level
    - Developer conducted a Cronbach's alpha (provider level: 0.94; facility level: 0.93)
    - Developer also estimated Spearman-Brown reliabilities (see table)
    - Data used in testing were gathered by the measure developer from the facilities described in its submission
- Reliability testing was performed at the measure score level

Validity

- Validity testing was performed at the data element and measure score levels.
- Validity of the measure score was established through face validity and empirical validity testing

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3543

Measure Title: Patient-Centered Contraceptive Counseling (PCCC) measure

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☒ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☒ **Composite**

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☒ **Assessment Data** ☐ **Paper Medical Records** ☒ **Instrument-Based Data** ☐ **Registry Data**
☐ **Enrollment Data** ☒ **Other**
Panel Member #3: Other: Team-based clinicians (See Table 2 of the testing document)
Panel Member #4: Patient survey responses, audio recordings of clinic visits

Level of Analysis:

☐ **Clinician: Group/Practice** ☒ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.** *[STAFF NOTE: The different text colors in this section are intended to differentiate the comments from each of the Methods Panel members.]*

   Panel Member #1: None.

   Panel Member #3: There appears to be some inconsistency in the required minimum number of responses for individual provider-level assessment (25 vs 30), and for facility-level assessment (30 or 50) between the 2 documents (S.15: Brief measure information vs. 2a2.3: Testing document). However, the latter provides the reliability estimates for range of panel sizes, and thus these numbers should guide future implementation.

   Panel Member #4: No Concerns

   Panel Member #5: None

   Panel Member #6: None.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**

   Panel Member #5: N/A
6. **Assess the method(s) used for reliability testing**

Submission document: Testing attachment, section 2a2.2

Panel Member #1: Data elements – calculated Cronbach's alpha

Measure score – signal-to-noise; appropriate method

Panel Member #2: Cronbach's alpha

Signal-to-noise ratio

Panel Member #3: Reliability testing methods are appropriately explained. Cronbach's alpha for data element reliability that measures internal consistency, while signal to noise ratio measured through Spearman-Brown measure of reliability were adopted.

Panel Member #4: Cronbach's alpha was used for the data element testing and intraclass correlation for the performance score. I have no concerns with the methods they used but I'm not sure about their method for scoring where each question has to be a perfect "5" to be score positive otherwise it was scored negative. Also, I would have like to see the reliability of the individual questions which I believe would have supported their method of all or none scoring.

Panel Member #5: I have no concerns related to reliability but am providing comments as requested. The measure developers address data element reliability by calculating Kronbach's alpha and address score-level reliability by estimating the proportion of total score-level variance explained by true signal variation as opposed to measurement error. Overall, I'm not sure the NQF requirement to estimate element-level reliability makes sense for this measure. Low element-level reliability for this measure is relevant to the extent that it impacts the score-level reliability but this is directly addressed in the developer's score-level reliability analysis and the score-level reliability is good. Kronbach's alpha measures internal consistency between items in the instrument and assesses the extent to which the number of items is sufficient in order for the average across items within an individual to be a reliable estimator of the individual's true underlying score on the instrument e.g. the estimate that would be obtained with an infinite number of questions. Kronbach's alpha assumes that instrument scores are calculated as averages across items in the instrument. Kronbach's alpha was arguably not applicable to the current measure because the items are not being averaged but being combined by assessing whether all of the items were in the highest category. Whereas Kronbach's alpha always increases with the number of items, the reliability of an assessment based on "all items were in the highest category" could decrease instead of increase as the number of items increases. Aside from the above issue, the methods and the description of the methods for assessing reliability were excellent. An implicit assumption is that the data distribution and signal variation in future populations using this measure will be similar to the sample used for reliability estimation.

Panel Member #6: The developer used Cronbach's alpha to evaluate reliability of the data elements and a signal to noise model using ICC. Both approaches are appropriate given that the PCCC is a composite measure, requiring evaluation of data element and measure score reliability.

7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3

Panel Member #1: Data elements –Cronbach's alpha values were good (0.93-0.94)

Measure score – Spearman-Brown Reliabilites (0.85 for facility; 0.84 for provider-level)

Panel Member #2: .94 and .93

Recommend panel size of 30 to achieve .84 with score

Panel Member #3: Cronbach's alpha for PCCC were 0.94 and 0.93 at provider- and facility-level, respectively. For the measure score, estimated provider-level reliability of 0.84 for 30 surveys and facility-level reliability of 0.85 for 50 surveys.

Panel Member #4: Statistical results for both the individual and performance score were acceptable.

Panel Member #5: For assessing individual providers, the developers estimate that 30 survey responses per provider will yield reliability of 0.84 (95% CI 0.67 to 0.93). For assessing facilities, the developers estimate that 50 survey responses per facility will yield reliability 0.85 (95% CI 0.74 to 0.92). This is good reliability. Achieving this volume in a reasonable timeframe would seem to be very feasible assuming response rates and volumes of contraceptive counseling patients are similar to those used in testing (Table 1)

Panel Member #6: Cronbach (item-level) and ICC (score-level) reliability scores indicate the measure is reliable at the data element level (measuring a single construct) for the compsite and PRO-PM aspects of the measure, and at the provider and practice level to evaluate whether the measure score is reasonably free from error.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    Panel Member #1: Tested reliability for both score and data elements; results were in the good to excellent range.

    Panel Member #2: Both elements and score following best practice.

    Panel Member #3: See responses in 6 and 7 above – all of the required issues have been addressed.

    Panel Member #4: Overall no major concerns but since this is a new measure based on a reduction of a larger survey instrument, I would have liked to see some analysis done around the reliability of the individual questions.

    Panel Member #5: No concerns. I rated high reliability because testing data demonstrates potential for high signal variation and the sample sizes required for 0.85 reliability seem to be achievable based on characteristics of providers in the testing data.

Panel Member #6: Overall, this measure appears to have good internal consistency among the PRO-PM data elements and is relatively free from measurement error at the provider/facility level.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

Submission document: Testing attachment, section 2b2.
Panel Member #1: None.

Panel Member #2: None

Panel Member #3: None

Panel Member #4: No concerns, exclusions appear appropriate.

Panel Member #5: None

Panel Member #6: None.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

Submission document: Testing attachment, section 2b4.

Panel Member #1: None. See variation in performance across measured units.

Panel Member #2: Range in performance, slightly right skewed, greater variation than CG-CAHPS.

Panel Member #3: None

Panel Member #4: No concerns

Panel Member #5: None

Panel Member #6: Provider and facility level performance is positively skewed, however, there is variation across both levels of measured entities.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

Submission document: Testing attachment, section 2b5.
Panel Member #1: Not applicable.
Panel Member #2: None

Panel Member #3: N/A
Panel Member #5: None

Panel Member #6: N/A

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

Panel Member #1: None.

Panel Member #2: None

Panel Member #3: None – all potential issues have been well-explained.

Panel Member #5: There is potential for selection bias if the probability of response differs depending on a patient's assessment of the provider. However, it seems possible that this bias would have a similar impact on all providers and would not be a major issue for making relative comparisons.

Panel Member #6: None.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☒ No ☐ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☒ Not applicable

Panel Member #5: N/A

16c.2 Conceptual rationale for social risk factors included? ☐ Yes ☒ No

Panel Member #5: N/A

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☒ No

Panel Member #5: N/A

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

Panel Member #5: N/A

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

Panel Member #5: N/A

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

Panel Member #5: N/A

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

Panel Member #5: N/A

16d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

Panel Member #5: N/A

16e**. Assess the risk-adjustment approach**

Panel Member #1: One thought…HCAHPS adjusts for patient-level factors, such as education level. Would it make sense to do something similar here?

Panel Member #4: Submitters note there is evidence that previous studies indicate race/ethnicity may impact the results but since the survey has not been tested for this they would consider it for a future analysis. Given that they had this data on hand, I believe they should have at least measured the differences and reported the results.

Panel Member #5: I don't object to the lack of risk adjustment but think the implications of this decision are worth noting. The developers argue that due to the question wording any differences across demographic groups would represent true differences in patient-centeredness. This argument seems to assume that individuals with a similar subjective experience or internal assessment will answer in the same Likert category. I am guessing this is not literally true due to possible person-specific differences in thresholds for assigning an excellent rating and possible

social & psychological influences e.g. desirability bias. Even if the assumption is true, an implication of not adjusting for case mix is that a provider who delivers less patient-centered care compared to another provider in all patient groups could still have a better measured performance score as a result of different case mix.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

    Panel Member #5: N/A

VALIDITY: TESTING

19. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☒ **Both**
20. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**

21. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    Panel Member #1: Face validity (measure score): assessed with two groups whether measure would differentiate quality
    Empirical testing (measure score): compared scores with two other PROs that address contraception
    Empirical testing (data element): compared responses to video capture of the encounter

    Panel Member #2: Validated observational approach

    Convergant validity, comparison with other items

    Face validity with hospital staff

    Panel Member #3: Critical data element validity was assessed through association between each of the 4 individual PCCC items with specific clinician communication practices consistent with patient-centered care, and assessed from audio recordings of visits using measures derived from Four Habits Coding Scheme (4HCS).
    Measure score validity was established through convergent validity assessment, which analyze the associations between the PCCC and two patient-reported measure using provider-level averages: (1) satisfaction with provider help with the choice of a birth control method, and (2) satisfaction with the method choice.
    Systematic assessment of face validity was conducted through input from facility administrators, providers of contraceptive counselling services, and patients (Modified Delphi process with administrators and providers; interview and focus groups with patients).

    Panel Member #4: Face validity very thorough. Validity of each question using the four habits coding scheme appears to be appropriate but I am not familiar with this method so I can't say if it is the correct method for validating this instrument.

    Panel Member #5: The developers addressed data element validity by assessing the associations between each individual PCCC item and assessments of communication practices made based on audio recording of encounters using a modified version of the Four Habits Coding Scheme (4HCS) framework for assessing and coding patient-

centered health communication. Methods of implementing the 4HCS were not described but presumably this was done by someone who was blinded to the patient's responses on PCCC items.

Score-level validity was assessed by comparing provider-level PCCC scores to provider-level scores calculated based on 2 measures of patient satisfaction.

In addition, a modified Delphi process was used to gather feedback from providers and administrators.

The methods were appropriate, and I have no major concerns about validity.

Panel Member #6: The developer measured evaluated validity in three ways: (1) evaluated the association between the PCCC items and use of practices consistent with patient centered care (2) evaluated PCCC score association with two satisfaction measures, and(3) conducted a systematic evaluation of the face validity. All approaches are reasonable and acceptable.

22. **Assess the results(s) for establishing validity**

Submission document: Testing attachment, section 2b2.3

Panel Member #1: Results for all three were very good (face validity: 85%+; score had expected relationship with other measures; all data elements were stat sig).

Panel Member #2: Data elements associated with relevant discussions

Related to other measures of quality

High ratings from facility staff

Panel Member #3: Detailed results provided in section 2b2.3 support the validity of the PCCC measure.

Panel Member #4: All analysis including the individual question and face validity strongly support that the survey instrument is valid.

Panel Member #5: Results for element validity included a set of regression coefficients describing the association between each single component of the PCCC instrument and a corresponding single component from the 4HCS framework. I was not sure of the exact interpretation of the regression coefficients because the developers did not specify (or I overlooked it) which of the 2 variables was the independent (explanatory) variable and which was the dependent (response) variable. Either way, it seems clear that the sets of variables were strongly correlated.

Regression coefficients for score-level validity demonstrated a strong association between provider-level performance on the proposed PCCC measure and provider-level scores for satisfaction about the choice of birth control method and satisfaction in how the provider helped the patient choose a birth control method.

Panel Member #6: Overall, the developer provided evidence that the items and the measure are valid.

23. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒ **Yes**

☐ **No**

☐ **Not applicable** (score-level testing was not performed)

24. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

Submission document: *Testing attachment, section 2b1.*

☒ **Yes**

☐ **No**

☐ **Not applicable** (data element testing was not performed)

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☒ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

    ☐ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

    Panel Member #1: Conducted testing on both score and data elements; methods were appropriate; results were robust.

    Panel Member #2: Multi-method, validated approach

    Panel Member #3: All the necessary components for demonstrating validity of the measure have been addressed.

    Panel Member #4: See above comments

    Panel Member #5: The measure directly assesses aspects of a patient's experience with a provider that have an undisputed relationship to patient-centered care and quality.

    Panel Member #6: Face validity was assessed appropriately, but represents a lower bar for validity testing. The empirical evaluation methods and findings were acceptable but are less compelling that correlating measure scores with a gold standard of patient centered care or another associated outcome of care. Regardless, validity testing results were acceptable.

ADDITIONAL RECOMMENDATIONS

27. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

    Panel Member #4: No, I don't believe my concerns were significant enough to keep this measure from proceed to the MAP committee

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability – Specifications

Comments:

**Data elements appear to be clearly defined, and given the short length of the survey, should be able to be implemented relatively easily (though this may be easier in clinics that already have satisfaction measures in place)

**None

**Cronbach's alpha is very impressive. Methodology panel approved the measure for reliability. Panel member 4 makes good points in asking whether testing data supports individual items and limiting credit to top-most scores on those items. Perhaps developers have further evaluation data to share.

**I wonder what the reading level of the questions are at to ensure that everyone will be able to read, process, and answer the questions reliably. Also, since this is a PRO, should there be a requirement to have information of who to contact if they have questions or concerns about their encounter?

**Involves asking each pt and then recording in a manner that can be extracted from the medical record. There will be significant resources need to administer and collate

**If the survey is administered as developed with no changes, it would alow for measure outcomes to have reliability. Concerns would be related to inconsistencies with components such as differences in the verbiage of the quesitons, differences in modality of survey, usinge different inclusion and exclusion criteria to administer survey, etc.

**In answer to the qustion of whether the measure con be consistently implemented, I am not sure how reliable this would be in a setting where a patient discusses birth control as part of a larger visit with other concerns. Her response may "lump" in her satisfaction with other aspects of the visit. In looking at the test sites, almost all of them were in family planning specific sites where the main mission is birth control counseling and provision.

**Elements were defined, not concerns for implementation

**Agree with the specifications and it is clear.

2a2. Reliability – Testing

Comments:

**No, though not sure I totally understand the calculations

**No

**No

**My only question is about the readability level.

**No

**No

**Cannot comment

**No concerns

**No

2b1. Validity –Testing

Comments:

**No.

**No

**No

**None

**No

**No

**Cannot comment

**No concerns

**No

2b2-3. Exclusions/Risk Adjustment

Comments:

**The exclusions appear appropriate, and I agree with excluding pregnant/postpartum women from this metric. They describe inclusion as those who are assigned female at birth, and am curious how this may be operationalized in a transgender male population

**Similar to the need to include women with low literacy, I question how women best served in a language other than English or Spanish, those who are blind, or those with intellectual impairment will be included in the survey responses.

**All ok, though very disappointing that we don't have a PRO-PM for use within maternity care.

**All items were met in the information presented. No concerns about the relationship between social risk factor variables and the measure focus.

**no risk adjustment needed

**none

**Risk adjustment was not tested. It may be that PRO survey results ARE sensitive to education level, cultural biases, or provider-patient racial concordancy.

**No concerns

**Providers will need to ensure that data elements for race and ethnicity are collected correctly. There is a great risk of inconsistencies in how race and ethnicity are determined at the various facilities/practices across the country.

2b4-7. Threats to Validity/Meaningful Differences/Comparability of Performance Scores/Missing Data

Comments:

**No concerns about face validity or missing data

**I am concerned that a percentage of women who declined to complete the survey did so because of low-literacy or illiteracy. These same women would have difficulty with printed contraceptive educational materials. Were instructions provided to practices to ensure that sub-groups of women were not excluded?

**No

**The measure appears to capture its intended data about the quality of the encounter with a provider specific to contraceptive care counseling.

**missing data, especially if missing data is different by race/ethnicity or langauge preference could threaten validity of the results

**None

**Again, I wonder whether these results would be different in a care delivery setting where women bring many concerns in during visits

**No

**No

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer reports that data for this measure must be collected through patient report. This measure cannot be captured through electronic data sources.
    - Note: for this criterion, “electronic data sources” means codes that can be automatically pulled (like claims data). The data for this measure for this must come directly from the patient so it’s

not considered as coming from electronic data sources, even if it's collected via a tablet (that is the mode of data collection for the tool).

- Facilities may opt to use either electronic or paper collection tools to collect responses. The developer does not require licensure or agreement from UCSF for use. There are no fees associated with use.
- The PCCC survey cannot be administered by the provider who gives the contraceptive counseling

Questions for the Committee:

- Is the data collection strategy ready to be put into operational use?
- Does the Committee have any concerns about the effort needed to administer the survey and calculate results on a national basis?
- Are there any concerns about feasibility from the patient perspective that have not yet been addressed?

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

**Contraceptive counseling is unevenly captured in the EHR, and clearly defining the denominator may be difficult- would it just be patients with a contraceptive ICD code? patients with a preventive care/health care maintenance code? used in contraception/family planning visits only?
**The materials described how providers and practices were "presented with final score reports" in the process of developing and testing this survey tool. However, once the survey is available for general use, practices will have to provide the resources to aggregate and report the survey responses. This will impose a cost on the practices and could be a barrier to implementation.
**This measure could be included in patient portal or Open Notes or Get My Health Data systems that interact with patients to streamline data collection..
**It is appropriate that the data cannot be extracted from the EHR. I do not have any concerns about this approach.
**New survey questions would need to be implemented and extracted
**Consistency in data collection would be important. Is there survey presented for all languages? If not, how is consistency is how the questions (when requiring translation) are asked validated?
**Post visit survey is a routine part of most care delivery systems. I anticipate difficulties with determining who is in the denominator seeking contraception. In many systems there may not be an appt type flag or problem reason code allowing identification of eligible patients. Also, if the strategy is on exiting the visit, how does an immediate response under the "eyes" of the care delivery system affect responses, compared to a mailed survey once a year or 4 weeks after the visit?
**Yes they should be available in EHR
**How will survey results be uploaded to the EMR or how will NQF obtain results?

Criterion 4: Usability and Use

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- Because this is a new measure, it is not yet in use. The developer is communicating with a number of entitles regarding implementation if the measure is endorsed.
- The developer hopes to incorporate the PCCC into requirements for PCMH certification within three years. In the long term, it plans to include it in the Uniform Data System and with OPA/Title X for public reporting within six years.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- All PCCC implementers were supplied with a report and interpretation of their results
- Developer also held meetings with staff and leaders at facilities. Developer used a Modified Delphi Process to elicit feedback from providers and administrators at facilities being measured
- Some facility leaders spoke about the "low feasibility of using EHR to identify patients to respond to the PCCC."
- None of the feedback received led to a change to measure specifications. However, the feedback received did lead to changes in implementation processes and the development of implementation guidelines.

Additional Feedback

N/A

Questions for the Committee:

- Can the performance results be used to further the goal of high-quality, efficient healthcare?
- Has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

- Because this is a new measure, the developer has not yet had an opportunity to collect improvement data, but suggests some opportunities for measuring improvement after implementation of the measure.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer found that implementation of the PCCC strengthened clinical workflows, specifically the check-out process.
- The developer states that patients also expressed appreciation for the ability to report their patient experience.
- The developer states that leadership at facilities being measured reported satisfaction with the tool as a way to gain insight into their patients

Potential harms

- The developer did not indicate any potential harms or unexpected benefits from this measure.

Additional Feedback:

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1-2. Use - Accountability and Transparency/Feedback

**Not currently being publicly reported or used. Received feedback from stakeholders via a modified Delphi process, which indicates some concerns around implementation

**I would like to see more information about how the PCCC measure might be used in publicly funded programs such as Title X and FQHCs.

**Public reporting and inclusion in accountability program are planned for this new, not-yet-endorsed measure. Endorsement would likely encourage inclusion in performance measurement programs. in the course of development and testing, results of implementation and feedback have been collected from women.

**Yes, this measure has been thoroughly discussed and vetted to ensure usability and appropriateness.

**I would encourage users to pair this tool with use of larc as originally envisioned by the devlopers. Large scale validation work looking at rates of unintended pregnancies would be ideal

**Measure not publicly reported or used in accountabilty application/program. From the developer, yes feedback has been provided to those being measured and they were allowed to provide feedback. the feedback was not used to change the measure however was used to changes processes and implementation guidelines.

**This measure is not proposed as being publically reported. If in the future this is proposed, testing would need a wider range of health delivery systems (not heavily weighted toward family planning settings). In general publically reported measures do not provide provider-specific data.

**Feedback

**The measure is not currently planned to be publically reported or in an accountability program. Feedback on the measure has been considered.

4b1. Usability – Improvement/ Benefits vs. harms/ Transparency

**Do not think there will be unintended harms, and the data presented appears to show some benefits for patients and clinics

**I believe that this measure can be valuable as a QA/QI tool for practices that have the resources to administer the survey, analyze the data, and report the results. However, many practices do not have the resources to perform these functions accurately and timely. In my experience, practices that perform patient satisfaction/experience surveys outsource the data analysis to a vendor. Are there systems in place to assist practices with this and how will practices secure the resources to accomplish this?

**This measure can discourage and/or document poor contraceptive counseling behavior and identify opportunities for improvement at both level of individual clinician and a care setting.

**For benefits vs. harm, I would encourage the developers to include sample language that users (hospitals/providers) can include about where someone could call, text, email to ask questions or express concerns about their experience talking to their provider about their contraceptive options.

**no harms to patients. Potential harms to practicioners if implemented incorrectly, provider could be mislabelled as a poor performer

**No potential harm or unintended benefits or consequences notes.

**I would be concerned about the problem of using this survey in conjunction with a very similar survey on member satisfaction with the office visit. Since a health system may have difficulty with determining who is coming in for birth control, it is possible for a patient to get double surveyed on their satisfaction with their visit

**I do not see any unintended consequences

**Understanding that the patient's perception of this care/interaction is planned to be evaluated could lead to improved perception of care delivery by the patient.

Criterion 5: Related and Competing Measures

Related or competing measures

- *Contraceptive Care – Most & Moderately Effective Methods* (NQF #2903)
- *Contraceptive Care – Access to LARC* (NQF #2904)
- *CG-CAHPS* (NQF #0005)

2903 and 2904 also address quality in the context of family planning care. 0005 is a general measure of patient experience and provider-patient communication. NQF staff do not assess these as competing measures.

Harmonization

N/A

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

**May be a useful complement to the existing contraceptive care measures

**No

**This measure complements and enhances but not compete with Contraceptive Care measures #2903 and #2904.

**None. This measure is needed.

**None known to me

**No competing measures that need to be harmonized.

**This harmonizes well with the LARC measures but may "Interfere" with other patient satisfaction surveys (see above)

**No

**None

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: January 21, 2020

- No comments have been submitted
- No NQF members have submitted a support/non-support choice:
  - 0 support the measure
  - 0 do not support the measure

Developer Submission

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

NQF_Evidence_attachment_UCSF_PCCC_7Nov2019_final.docx

1a.*1 For Maintenance of Endorsement:* Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a. Evidence (subcriterion 1a)

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** #3543 (not yet endorsed)

Measure Title: The Patient-Centered Contraceptive Counseling (PCCC) measure

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 11/8/219

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome: Patient experience of contraceptive counseling

☒ Patient-reported outcome (PRO): Patient experience of contraceptive counseling

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (e.g., lab value):

☐ Process:

☐ Appropriate use measure:

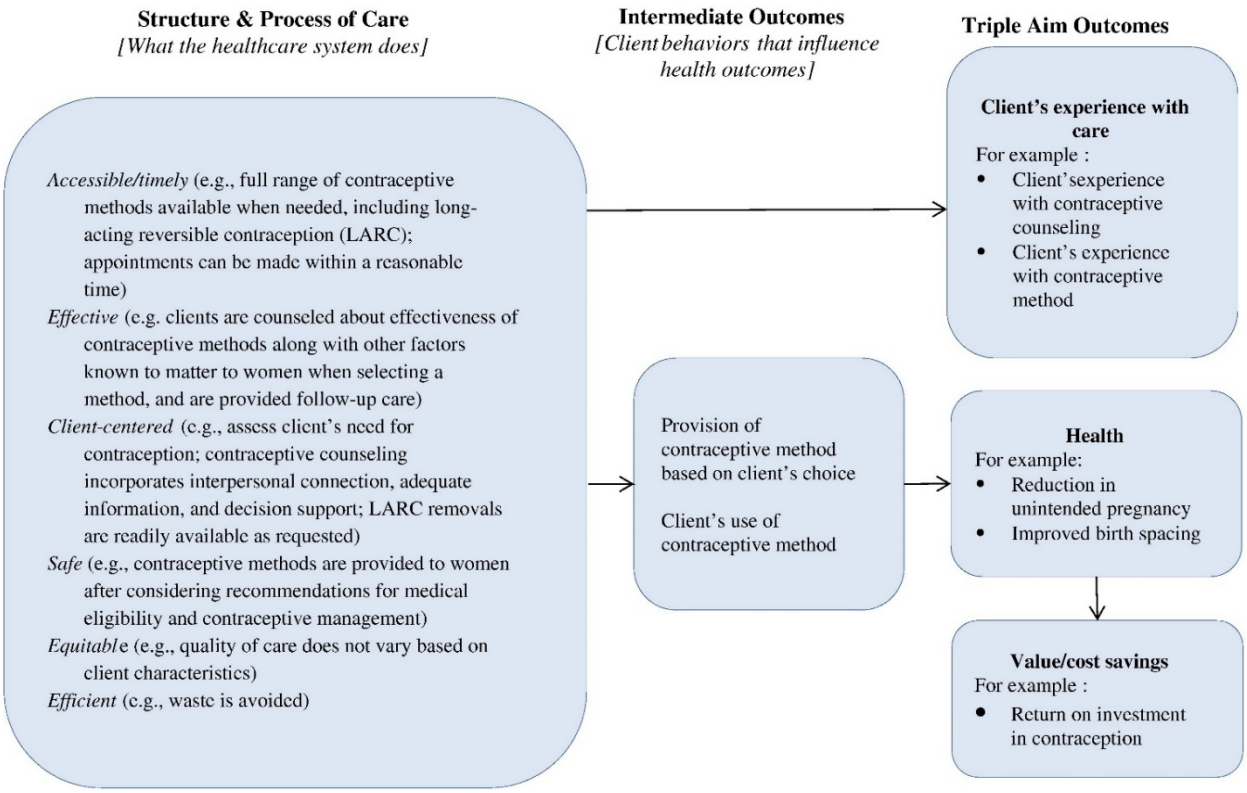
☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The diagram in Figure 1 below describes the relationship between the structures and processes of quality contraceptive care, including patient- (or client-) centered care, and improved outcomes, including the outcome of relevance for this application, patient/client experience. This diagram was developed in 2017 by the Department of Health and Human Services, Office of Population Affairs (OPA), in collaboration with Christine Dehlendorf, Principal Investigator on the development and testing of the PCCC. The diagram was presented in the context of describing OPA's work to develop claims-based measures of contraceptive provision (endorsed by NQF in 2016), and the need for the development of a Patient-Reported Outcome Performance Measure (PRO-PM) to help provide a more robust picture of contraceptive care quality beyond the currently endorsed measures.[1]

Figure 1. Office of Population Affairs' conceptual model for clinical performance measures for contraceptive care.



1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

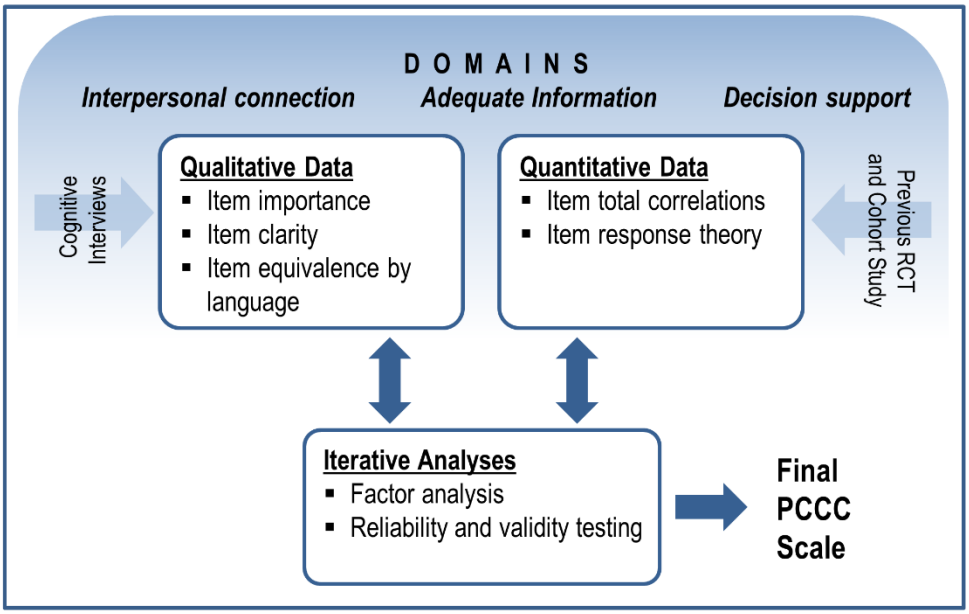
Throughout the development and testing the PCCC as a measure of patient experience, we have put substantial effort toward ensuring that this measure reflects the values and preferences of patients themselves with respect to their experience of counseling.

The development of the PCCC was from the outset informed by a 2009 qualitative study by our team, in which we conducted 42 in-depth interviews with English- and Spanish-speaking patients about their preferences for contraceptive counseling.[2] We recruited patients following contraceptive counseling visits at five publicly funded San Francisco Bay Area clinics. Participants were 24% Black Non-Hispanic/Latina, 24% White Non-Hispanic/Latina, and 52% Hispanic/Latina. Thirteen interviews were conducted in Spanish. We used a modified grounded theory methodology to assess emergent qualitative themes in interview transcripts. Patients reported that having positive experiences with contraceptive counseling was highly valuable, particularly due to the unique and sensitive nature of decisions around sex and pregnancy. Through this work, we identified three domains of patient-centered contraceptive counseling valued by patients: interpersonal connection between provider and patient, adequate information, and decision support. Our measure development work has been guided by these three domains, with regular input from patients themselves to ensure that the measure continues to resonate with patients.

Drawing on the three domains of patient-centered contraceptive counseling derived from our qualitative work, we developed an the 11-item Interpersonal Quality of Family Planning care (IQFP) Patient-Reported Outcome Measure (PROM) for use in the research context. To develop the IQFP, our team drew on the findings of our previous qualitative research and adapted 17 items total from the Consultation and Relationship Empathy (CARE) scale[3] and the Interpersonal Process of Care (IPC) scale[4] to the context of contraceptive care. We assessed interim correlations and performed exploratory factor analysis using data from a cohort study of 346 women who had received contraceptive counseling in order to select 11 items for the IQFP, which together best captured the experience of patient-centered contraceptive counseling.[5] We demonstrated that the English-language version of the IQFP has content, construct, convergent and discriminant validity,[5] as well as predictive validity with the outcome of contraceptive continuation at six months.[6]

To develop a PRO-PM, we identified the need to reduce the number of items from the IQFP in order to enhance real-world feasibility, as well as to ensure language equivalence in both English and Spanish. In order to accomplish this, we again worked with patients to understand their preferences and priorities for contraceptive counseling. We used an iterative data triangulation process (Figure 2) drawing on qualitative data from patient interviews and quantitative data from previous patient responses to the IQFP (n=1,097) to select items for a reduced measure that prioritized patient feedback, retained the validity of the IQFP in psychometric testing, and represented all three domains of patient-centeredness in contraceptive counseling. This process was described briefly in Section 2a2.3 of our testing attachment and is described in more depth below.

Figure 2. Qualitative and quantitative data triangulation for item reduction of the initial IQFP to the PCCC



In order to select items for the reduced measure, we administered the IQFP and conducted English and Spanish cognitive interviews with patients who had recently received contraceptive counseling (n=33) at three publicly-funded California health facilities, exploring item importance, clarity, and language equivalence to patients in order to identify methods to prioritize for inclusion in the reduced scale. Participants included ten monolingual English speakers, thirteen monolingual Spanish speakers, and ten bilingual English and Spanish speakers. We evaluated items for equivalence between English and Spanish versions both by comparing interpretations of item meaning by

English and Spanish speakers, and asking bilingual participants to compare the meanings of English and Spanish versions themselves.

We used content analysis to understand patient responses to each item and their overall experience of taking the survey. We also calculated mean rankings of items to quantitatively understand ranked item importance. Drawing on this data, we performed psychometric testing on combinations of three to six IQFP items that could serve as a reduced measure using previously collected measure data. Internal consistency of reduced item combinations was tested using Cronbach's Alpha. Factor loadings were used to test for consistency of a single factor analysis solution. External validity was tested by examining the association of item combination responses with other items included in post-visit patient surveys. These included measures of global satisfaction with the visit (concurrent validity), satisfaction with the method selection process (concurrent validity), satisfaction with method choice (convergent validity), and recommendation of the provider to a friend (convergent validity). Predictive validity was also tested by examining the association between item combination responses and follow-up data of contraceptive continuation at six months, which we had obtained from our previous study using the IQFP measure. Cronbach's alpha and measures of external validity for each item combination were compared to findings that validated the original IQFP, to understand how well each combination retained the original's properties. As the IQFP was validated as a dichotomous score (top score of 55 versus less than 55), we calculated the sensitivity and specificity of top scores of reduced item combinations for a top IQFP score.

The resulting PCCC includes four items that were found in both qualitative and quantitative analyses to retain the IQFP's ability to capture patient-centeredness in contraceptive counseling. These items were understandable to patients and equitable across languages. Patients also ranked PCCC items relatively highly in terms of importance, as compared to other items in their domain. Psychometric analysis results for the PCCC as compared with the 11-item IQFP are depicted in Table 1 below.

Table 1. Psychometric Results for the IQFP and PCCC

Property	Test	Result for IQFP	Result for PCCC
Internal consistency	Cronbach's alpha	0.97	0.92
Concurrent validity	Association with measure of high global satisfaction with visit	$p<0.001$	$p<0.001$
Concurrent validity	Association with measure of satisfaction with method selection process	$p<0.001$	$p<0.001$
Convergent validity	Association with measure of satisfaction with method choice	$p<0.01$	$p<0.01$

Convergent validity	Association of measure of recommendation of provider to a friend	p<0.01	p<0.01
Predictive validity	Association with contraceptive continuation at 6 months	p<0.05	p<0.05

The association between a top-box response on the PCCC and satisfaction with choice of method has also been observed in a representative sample of 1,234 women of reproductive age residing in the U.S. South (unpublished data, personal communication, Khoury A, August 28, 2019).

To further ensure meaningfulness of the 4-item PCCC before piloting it with health facilities across the country, we gathered feedback on the face validity of the reduced measure in interviews and focus groups with 43 patients from California, Texas, and North Carolina. This research is described in Section 2b1.2 of our testing attachment to this application. As noted there, examination of face validity included questions about the acceptability of the measure itself and the utility of a measure score to patients. Interviews and focus groups revealed that patients found the items and performance measure score were acceptable and useful to them. Eighty-eight percent of patients reported that a facility or provider having a higher score on the performance measure would make them more likely to choose that facility or provider for their care, as opposed to having no effect on decision making about their care.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

As depicted in Figure 1 of Section 1a.2, the outcome of patient experience in contraceptive counseling, as measured by the PCCC, has inherent value as an outcome in and of itself. This outcome is valuable and meaningful to patients, as described in Section 1a.3. Below we describe evidence supporting the ability of the health care system to act to improve scores on this measure. We conclude with data of relevance to the impact of this PRO-PM on clinical outcomes.

The PCCC focuses explicitly on the counseling interaction between providers and patients. Processes or interventions influencing this interaction therefore have the potential to influence scores on this measure. One such evidence-based tool is the *My Birth Control* decision support tool for contraceptive counseling, which uses a shared decision-making approach to help center patient preferences for their birth control in contraceptive counseling conversations.[7] Using data from a randomized controlled trial of 749 women in the San Francisco Bay Area, we found that use of *My Birth Control* was associated with a top-box response to the PCCC (with 72.4% of

intervention-arm participants giving a top-box response on the PCCC vs. 64.7% of control-arm participants, p=0.026 [unpublished data]).

More generally, evidence supports the proposition that health care system interventions, including, training and counseling interventions targeted at patient-provider communication can produce change in patient experience outcomes. These include studies of decision aids in other areas of health care, which have found a beneficial impact on patient experience.[8] The use of CAHPS surveys provides examples of PRO-PMs related to patient communication being responsive to interventions. In the hospital setting, HCAHPS responses related to communication have been shown to improve following educational interventions on quality patient communication for hospital providers and staff[9] and for nurses specifically.[10] Similar findings have been found for use of other PRO-PMs such as Press-Ganey surveys.[11] Upon its wider dissemination, we envision the PCCC as a readily available tool that can be used to help develop and evaluate quality improvement efforts specific to the contraceptive counseling interaction.

In addition to providing important information for health care organizations on patient experience itself, the PCCC is also associated with contraceptive continuation. Like the IQFP,[6] the PCCC has predictive validity with the outcome of contraceptive continuation at six months follow-up (see Table 1 in Section 1a.3 above). By working to improve patient experience, health care organizations can therefore help support their patients achieve their reproductive goals, such as pregnancy prevention. Further, qualitative data suggests that patients who experience non-patient-centered care are less likely to return to seek out care for future reproductive health needs.[12] This has the potential to negatively impact a range of outcomes, including pregnancy-related morbidity and mortality. An association between patient experience and reproductive health outcomes is aligned with a larger body of evidence of how patient experience is associated with a range of health outcomes, structures, and processes. As described in the CAHPS Clinician and Group Survey Version 3.0 application, there is a substantial body of literature documenting evidence of the relationships between patient experience, clinical processes, and patient outcomes.[13] This includes two systematic reviews (one UK-based[14] and one US-based[15]), which both found that patient experience was positively associated with outcomes such as seeking and adhering to preventive care treatments, and positive clinical health outcomes. In addition, a 2009 meta-analysis of 127 studies documented that the quality of provider communication – which is the specific aspect of patient-centeredness measured by the PCCC – was directly associated with patient treatment adherence.[16]

The PCCC is therefore an actionable measure that allows health care organizations, facilities, and providers the opportunity to understand their patients' experience of contraceptive counseling and implement quality improvement strategies to improve patient experience and clinical outcomes.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

References

1. Gavin LE, Ahrens KA, Dehlendorf C, Frederiksen BN, Decker E, Moskosky S. Future directions in performance measures for contraceptive care: a proposed framework. *Contraception.* 2017;96(3):138-144.
2. Dehlendorf C, Levy K, Kelley A, Grumbach K, Steinauer J. Women's preferences for contraceptive counseling and decision making. *Contraception.* 2013;88(2):250-256.
3. Mercer SW, McConnachie A, Maxwell M, Heaney D, Watt GC. Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Family Practice.* 2005;22(3):328-334.
4. Stewart AL, Nápoles-Springer AM, Gregorich SE, Santoyo-Olsson JJ. Interpersonal processes of care survey: Patient-reported measures for diverse groups. *Health Services Research.* 2007;42(3p1):1235-1256.
5. Dehlendorf C, Henderson JT, Vittinghoff E, Steinauer J, Hessler DJC. Development of a patient-reported measure of the interpersonal quality of family planning care. *Contraception.* 2018;97(1):34-40.
6. Dehlendorf C, Henderson JT, Vittinghoff E, et al. Association of the quality of interpersonal care during family planning counseling with contraceptive use. *American Journal of Obstetrics and Gynecology.* 2016;215(1):78. e71-78. e79.
7. Dehlendorf C, Fitzpatrick J, Steinauer J, et al. Development and field testing of a decision support tool to facilitate shared decision making in contraceptive counseling. *Patient Education and Counseling.* 2017;100(7):1374-1381.
8. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews.* 2017;4(4):CD001431-CD001431.
9. Horton DJ, Yarbrough PM, Wanner N, Murphy RD, Kukhareva PV, Kawamoto KJ. Improving physician communication with patients as measured by HCAHPS using a standardized communication model. *American Journal of Medical Quality.* 2017;32(6):617-624.
10. Briggs K, Sharma L, Chandrasekaran A, Douglas C, Aroh D, Finefrock D. The effect of a hybrid training program: improving nursing communication skills and HCAHPS scores. *Nursing Management.* 2018;49(2):51-53.
11. Fustino NJ, Kochanski JJ. Improving Patient Satisfaction in a Midsize Pediatric Hematology-Oncology Outpatient Clinic. *Journal of Oncology Practice.* 2015;11(5):416-420.
12. Gomez AM, Wapman M. Under (implicit) pressure: young Black and Latina women's perceptions of contraceptive care. *Contraception.* 2017;96(4):221-226.
13. Diedrich J, Steinauer J. Complications of surgical abortion. *Clinical Obstetrics and Gynecology.* 2009;52(2):205-212.
14. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. *BMJ Open.* 2013;3(1):e001570.
15. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. *Medical Care Research and Review.* 2014;71(5):522-554.

16. Zolnierek KBH, DiMatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. *Medical Care.* 2009;47(8):826.

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

The PCCC is designed to give health care organizations, facilities, and providers the opportunity to measure the quality of their patients' experience of contraceptive counseling and implement quality improvement strategies to improve patient experience as needed. While PCCC results are intended to have stand-alone value to organizations, we also intend for the PCCC to serve as a balancing measure for currently endorsed measures of contraceptive provision (NQF measures #2903 and #2904), of which the Office of Population Affairs (OPA) is the steward. Below, we describe the rationales for a measure of patient experience of contraceptive counseling in its own right, and the rationale for this measure's use alongside contraceptive provision measures.

Rationale for a measure of patient experience of contraceptive counseling

As described in our Evidence Attachment, patient experience of contraceptive counseling is an important outcome in and of itself, in that it is highly valued by patients [1] and measures a core aspect of quality care – patient-centeredness – as defined by the Institute of Medicine in its report Crossing the Quality Chasm [2]. This is consistent with the National Quality Forum's consideration of patient experience as one of the domains of Patient Reported Outcomes, as described in the measure evaluation criteria guidance document [3]. In order to capture this outcome with a PRO-PM, we engaged in a process of measure development that was continually informed by the input of patients on their needs and preferences for this type of care. Therefore, the resulting PCCC measure allows for identification of whether patients are experiencing high quality care as they themselves define it. In addition, patient-centeredness of contraceptive counseling as measured by the PCCC has been demonstrated to be associated with contraceptive continuation at six months [4], indicating a relationship between patient experience of counseling and the ability of patients to achieve their own reproductive goals, including pregnancy prevention. Patient experience has also been linked to improved engagement with care in various contexts [5,6]; in the context of contraceptive care, this means that patients who receive patient-centered care may feel more able to continue engaging with the reproductive health care system not only for contraception, but also if and when they become pregnant and/or give birth [7]. As such, positive patient experience of contraceptive counseling can support positive pregnancy and birth outcomes such as reduced maternal mortality.

Given the important implications of patient-centeredness of contraceptive counseling, both for patient experience and reproductive health outcomes, many health care organizations are understandably invested in gathering information on the experiences of their patients and improving those experiences as needed. The PCCC serves as a tool that organizations can use to understand the patient-centeredness of counseling and evaluate quality improvement interventions. CAHPS measures have been used to monitor the effectiveness of educational interventions for providers and staff to improve patient experience [8,9]. Similarly, the PCCC may be used to inform quality improvement activities, such as contraceptive counseling training or implementation of decision support tools to support contraceptive decision making, and to track their impact over time [10].

Rationale for use alongside contraceptive provision measures

The motivation behind the development of the PCCC originated during OPA's development of measures #2903 and #2904, which focus on most and moderately effective (MME) contraception and long-acting reversible contraceptive (LARC) methods. The OPA team and others involved in the measure development process foresaw that use of these important measures could have the unintended consequence of incentivizing provider pressure on patients to use more effective methods. During the NQF endorsement process, this concern was voiced by stakeholders, including the National Partnership for Women & Families (NPWF). The NPWF submitted a public comment that stated, "It is extremely important to keep in mind that reproductive coercion has a troubling history, and remains an ongoing reality for many, including low-income women, women of color, young women, immigrant women, LGBT people, and incarcerated women. We hope this measure will be paired with a woman-reported 'balancing measure' of experience of receiving contraceptive care. Such a measure can be expected to help identify and/or check inappropriate pressure from the health care system." Following endorsement of the contraceptive provision measures by the NQF, OPA acted on this shared concern by funding a three-year cooperative agreement with UCSF in order to develop a PRO-PM as a 'balancing measure' to support proper use of the provision measures, and to enable health facilities and systems to measure patient experience in its own right. This initial funding supported our team at UCSF to work to reduce the IQFP to become the PCCC, as described in our Evidence Attachment. Further private foundation funding has supported the PCCC's validity and reliability testing in health care settings across the country.

Our team at UCSF intends to conduct further work to optimize use of the provision measures and the PCCC together. We hypothesize that the PCCC will serve as balancing measure for the provision measures, so that organizations can observe any fluctuations in PCCC scores that occur in association with changes in provision scores, and ensure that any increased provision does not come at the cost of patient experience, or ideally would in fact be associated with improved patient experience. As such, use of these two types of measures together can result in a more robust picture of contraceptive care quality, and assist health care organizations to achieve both aspects of quality in contraceptive care: providing access to a range of contraceptive methods and providing patient-centered counseling free of coercion.

References

[1] Dehlendorf C, Levy K, Kelley A, Grumbach K, Steinauer J. Women´s preferences for contraceptive counseling and decision making. Contraception. 2013;88(2):250-256.

[2] Wolfe A. Institute of Medicine Report: crossing the quality chasm: a new health care system for the 21st century. Policy, Politics, & Nursing Practice. 2001;2(3):233-235.

[3] National Quality Forum. Measure evaluation criteria and guidance for evaluating measures for endorsement. 2018. http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/2018_Measure_Evaluation_Criteria_and_Guidance.aspx. Accessed 31 Jul 2019.

[4] Dehlendorf C, Henderson JT, Vittinghoff E, et al. Association of the quality of interpersonal care during family planning counseling with contraceptive use. American Journal of Obstetrics and Gynecology. 2016;215(1):78. e71-78. e79.

[5] Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. Medical Care Research and Review. 2014;71(5):522-554.

[6] Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open. 2013;3(1):e001570.

[7] Gomez AM, Wapman M. Under (implicit) pressure: young Black and Latina women´s perceptions of contraceptive care. Contraception. 2017;96(4):221-226.

[8] Horton DJ, Yarbrough PM, Wanner N, Murphy RD, Kukhareva PV, Kawamoto K. Improving physician communication with patients as measured by HCAHPS using a standardized communication model. American Journal of Medical Quality. 2017;32(6):617-624.

[9] Briggs K, Sharma L, Chandrasekaran A, Douglas C, Aroh D, Finefrock D. The effect of a hybrid training program. Nursing Management. 2018;49(2):51-53.

[10] Dehlendorf C, Fitzpatrick J, Fox E, et al. Cluster randomized trial of a patient-centered contraceptive decision support tool, My Birth Control. American Journal of Obstetrics and Gynecology. 2019;220(6):565. e561-565. e512.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. ***(This is required for maintenance of endorsement.*** *Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

As described in our Testing Attachment, we have analyzed the PCCC as a measure of patient experience of contraceptive counseling at both the individual provider level and health care facility level. Patient responses were collected at each level of analysis at a single time-point in each participating facility (data were not collected over time for comparison at different time-points). The 34 participating providers were employed at ten of the participating facilities. The remaining 12 participating facilities did not submit provider-level data. We include descriptive statistics on each level of analysis below. Please refer to the Testing Attachment for complete descriptions of participating entities and histograms depicting the distributions of scores at both levels of analysis.

- Provider-level analysis (n=34 providers who provided counseling to 2,477 patients)
- Mean performance score: 0.81
- Standard deviation: 0.12
- Range: 0.44-0.95
- Percentiles
- 25th: 0.79
- 50th: 0.85
- 75th: 0.90
- Facility-level analysis (n=22 facilities that provided counseling to 3,478 patients)
- Mean: 0.79
- Standard deviation: 0.12
- Range: 0.51-0.97
- Percentiles
- 25th: 0.70
- 50th: 0.83
- 75th: 0.88

These results indicate that while the performance measure scores are left-skewed in a manner similar to many patient satisfaction and experience measures [1], there is variability in scores indicating the opportunity for improvement among low performers. Of note, our distribution is wider than that of available statistics on the CG-CAHPS communication composite score, in which the median score at a clinic level is 88%, with a 25th/75th percentile of 84%/91% [2].

References

[1] Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Social Science & Medicine. 1998;47(9):1351-1359.

[2] Agency for Healthcare Research and Quality. CAHPS aggregated data. https://cahpsdatabase.ahrq.gov/CAHPSIDB/CG/Percentile.aspx. Accessed June 30, 2019.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

As depicted in 1b.2 above, data from our reliability and validity testing indicate a range in provider and facility performance on the PCCC. These findings are consistent with evidence in the literature indicating substantial room for improvement in the patient-centeredness of contraceptive counseling in the United States. In multiple studies examining patient experience of counseling, patients reported receiving inadequate information from their providers to make an informed decision on contraception [1-4], and feeling dissatisfied with the patient-centeredness and adequacy of counseling overall [4-7]. Furthermore, research conducted by our team at UCSF examining quality of counseling via audio recording of patient visits found that providers inconsistently elicited or engaged with patient experiences and preferences during counseling [8,9].

References

[1] Dehlendorf C, Levy K, Kelley A, Grumbach K, Steinauer J. Women´s preferences for contraceptive counseling and decision making. Contraception. 2013;88(2):250-256.

[2] Yee LM, Simon MA. Perceptions of coercion, discrimination and other negative experiences in postpartum contraceptive counseling for low-income minority women. Journal of Health Care for the Poor and Underserved. 2011;22(4):1387-1400.

[3] Guendelman S, Denny C, Mauldon J, Chetkovich C. Perceptions of hormonal contraceptive safety and side effects among low-income Latina and non-Latina women. Maternal and Child Health Journal. 2000;4(4):233-239.

[4] Becker D, Koenig MA, Mi Kim Y, Cardona K, Sonenstein F. The quality of family planning services in the United States: findings from a literature review. Perspectives on Sexual and Reproductive Health. 2007;39(4):206-215.

[5] Becker D, Tsui AOJPoS, Health R. Reproductive health service preferences and perceptions of quality among low-income women: racial, ethnic and language group differences. Perspectives on Sexual and Reproductive Health. 2008;40(4):202-211.

[6] Nobili MP, Piergrossi S, Brusati V, Moja E. The effect of patient-centered contraceptive counseling in women who undergo a voluntary termination of pregnancy. Patient Education and Counseling. 2007;65(3):361-368.

[7] Borrero S, Schwarz EB, Creinin M, Ibrahim S. The impact of race and ethnicity on receipt of family planning services in the United States. Journal of Women´s Health. 2009;18(1):91-96.

[8] Dehlendorf C, Anderson N, Vittinghoff E, Grumbach K, Levy K, Steinauer J. Quality and content of patient–provider communication about contraception: differences by race/ethnicity and socioeconomic status. Women´s Health Issues. 2017;27(5):530-538.

[9] Dehlendorf C, Kimport K, Levy K, Steinauer J. A qualitative analysis of approaches to contraceptive counseling. Perspectives on Sexual and Reproductive Health. 2014;46(4):233-240.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. ***(This is required for maintenance of endorsement****. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We analyzed patient response to the PCCC by self-reported race and ethnicity, and language in which they completed the PCCC in the facility-level sample used in reliability testing, as evidence suggests that patient experience can differ by racial, ethnic, and language categories (see Section 1b.5 below). Analysis of responses by race included the 2,820 patients for whom we had data on race. Analysis of responses by ethnicity included the 3,378 patients who reported their ethnicity. See testing attachment, section 1.6 for an explanation of different sample size for these three types of demographics. Patients had visits to 22 different facilities.

In our overall sample, people of color less frequently gave a top-box score on the PCCC as compared with White respondents (with an overall percentage of top-box score of 78.5% for Black patients and 74.4% for Asian patients, compared with 86.0% for White patients, and 80.4% for Hispanic/Latina patients compared with 84.1% for non-Hispanic/Latina patients). Looking across facilities, we identified a total of 9 facilities in which White participants had more than a 5% higher percentage in top-box scores compared with other patients, 5 facilities in which White patients had higher percentages not greater than 5%, and 7 facilities in which non-White patients had higher percentages. (In one facility, no patients identified as White, and thus no such comparison could be drawn.) Similarly, by ethnicity, in 7 facilities non-Hispanic/Latina patients had more than a 5% higher percentage of top-box scores than Hispanic/Latina patients, 4 facilities had a less than 5% difference, and 10 had the opposite pattern. (In one facility, all patients identified as Hispanic/Latina.) This variability in disparities across facilities underscores the opportunity for facilities with substantial gaps in performance between racial and ethnic groups to conduct quality improvement in a manner that enhances equity.

Spanish speakers less frequently gave a top-box score compared with English speakers in our overall sample (83.9% vs. 68.2% overall across facilities). We consider these results to reflect actual, meaningful differences in care, rather than different item interpretation or response patterns based on cultural or linguistic differences. This conclusion is based on the evidence for differential quality of care from existing literature, as described in Section 1b.5, as well as our rigorous process of developing the Spanish language measure alongside the English language measure. This process included ensuring equivalence of items across language in our cognitive interviews, as well as collecting data on item importance from Spanish and English speakers, and using this information to determine the items included in our measure, as well as conducting face validity testing with both Spanish and English speakers. This process is described in more detail in our evidence attachment, Section 1a.3. Similar to our findings by race and ethnicity, there was variability in disparities by language across facilities. In 5 of the 9 facilities with more than 5 Spanish-speaking respondents, English speakers had more than a 5% higher percentage of top-box scores compared with Spanish speakers; in one facility; English speakers had a less than 5% greater percentage; and in 3 facilities, the reverse pattern emerged. Of note, among the sites in which the reverse pattern was true is a site with which our team has worked closely in the past. We have become well acquainted with the approaches they use to optimize care for Hispanic/Latina and Spanish-speaking patients, including having clinical staff who are both culturally- and language-concordant with this population. Given this, when first considering our findings of differences by language, and prior to identifying individual facility scores, we hypothesized that this facility would not have a lower percentage of top-box scores for the PCCC among Spanish speakers, and that if true, this would provide further support for the overall lower scores in this population reflecting meaningful differences. This was in fact the case, as we found a higher score of 85% among Spanish speaking patients, compared to 65% among English speaking patients, from this facility.

For tables of results, please see Appendix – PCCC response by race, ethnicity, and language.

As noted in our testing attachment, we consider it critical to give attention to how patient race, ethnicity, and language may affect provider-patient communication and patient experience of care, and to consider the possibility of stratified reporting of results in the future. We intend to explore such differences in depth in further analyses of the PCCC and will report these as appropriate during maintenance of endorsement.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

In addition to extensive data documenting differences in the quality of health communication by race/ethnicity across areas of health care [1-5], research in contraceptive services specifically have found that patients of color report receiving lower quality of contraceptive services as compared with White patients [6-8]. Importantly, Black and Latina patients have reported perceiving bias among providers of contraceptive counseling based on patient background, with providers recommending specific methods over others in a non-patient-centered manner [9-11]. In a recent qualitative study using in-depth interviews with Black and Latina patients, participants reported non-patient-centeredness in counseling manifesting in imbalanced provision of information – including failure to share adequate information about side effects – and tone of voice indicating provider preference for more effective methods [9]. These findings evoke concerns voiced by the NPWF and others during the comment period for measures of contraceptive provision that an emphasis on method efficacy can lead to coercive counseling, particularly for patients of color, and that a PRO-PM for patient experience of counseling can help mitigate any such effects.

With regard to language, Spanish-speaking patients with limited English proficiency experience lower quality of care compared with fluent English speakers due to language barriers disrupting the ease of communication between provider and patient [12]. This difference is especially evident in language-discordant interactions, where the provider is not Spanish-speaking [13]. The quality of interpretation services is highly important in mitigating the burden of language-discordance, and these services are not universally available. In a 2002 study at a walk-in clinic in Denver, Spanish-speaking patients who used family members or ad hoc interpreters were less satisfied with provider listening, discussion of sensitive issues, explanations, answers, and support compared with language-concordant counterparts and those who used professional interpreter services [14]. Spanish speakers have reported the importance of language-appropriate contraceptive care services in particular, with reliance on interpreter services for contraceptive counseling leading to patient embarrassment in sharing personal information with an interpreter [6].

[1] Aseltine Jr RH, Sabina A, Barclay G, Graham G. Variation in patient–provider communication by patient's race and ethnicity, provider type, and continuity in and site of care: An analysis of data from the Connecticut Health Care Survey. SAGE Open Medicine. 2016;4:2050312115625162.

[2] Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, Hays R. Race/ethnicity, language, and patients´ assessments of care in Medicaid managed care. Health Services Research. Health Services Research. 2003;38(3):789-808.

[3] Maina IW, Belton TD, Ginzberg S, Singh A, Johnson T. A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. Social Science & Medicine. 2018;199:219-229.

[4] Palmer NR, Kent EE, Forsythe LP, et al. Racial and ethnic disparities in patient-provider communication, quality-of-care ratings, and patient activation among long-term cancer survivors. Journal of Clinical Oncology. 2014;32(36):4087.

[5] Dahlem CHY, Villarruel AM, Ronis DL. African American women and prenatal care: perceptions of patient–provider interaction. Western Journal of Nursing Research. 2015;37(2):217-235.

[6] Becker D, Tsui AOJPoS, Health R. Reproductive health service preferences and perceptions of quality among low-income women: racial, ethnic and language group differences. Perspectives on Sexual and Reproductive Health. 2008;40(4):202-211.

[7] Downing RA, LaVeist TA, Bullock HE. Intersections of ethnicity and social class in provider advice regarding reproductive health. American Journal of Public Health. 2007;97(10):1803-1807.

[8] Forrest JD, Frost JJ. The family planning attitudes and experiences of low-income women. Family Planning Perspectives. 1996:246-277.

[9] Gomez AM, Wapman M. Under (implicit) pressure: young Black and Latina women´s perceptions of contraceptive care. Contraception. 2017;96(4):221-226.

[10] Yee LM, Simon MA. Perceptions of coercion, discrimination and other negative experiences in postpartum contraceptive counseling for low-income minority women. Journal of Health Care for the Poor and Underserved. 2011;22(4):1387-1400.

[11] Borrero S, Schwarz EB, Creinin M, Ibrahim S. The impact of race and ethnicity on receipt of family planning services in the United States. Journal of Women´s Health. 2009;18(1):91-96.

[12] Doty MM. Hispanic patients´ double burden: Lack of health insurance and limited English. Commonwealth Fund New York; 2003.

[13] Sudore RL, Landefeld CS, Perez-Stable EJ, et al. Unraveling the relationship between literacy, language proficiency, and patient–physician communication. Patient Education and Counseling. 2009;75(3):398-402.

[14] Lee LJ, Batal HA, Maselli JH, Kutner JS. Effect of Spanish interpretation method on patient satisfaction in an urban walk-in clinic. Journal of General Internal Medicine. 2002;17(8):641-646.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

https://pcccmeasure.ucsf.edu/

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

No data dictionary **Attachment:**

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Attachment **Attachment:** PCCC_instrument.docx

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Patient

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

N/A

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, *state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

The PCCC is a visit-specific measure of patient-centeredness in contraceptive counseling. It specifically measures how many patients report a top-box (i.e., the highest possible) score of patient experience in their contraceptive counseling interaction with a health care provider during their recent visit.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, *describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

Identification in the numerator is determined by patient response to the PCCC. The numerator for both the individual provider and facility level includes only those patients who gave a top-box score for their interaction with their health care provider on the PCCC. All other conditions determining inclusion in the numerator also determine inclusion in the denominator. As such, please see response to S.7. for additional details on inclusion.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The target population for the PCCC is patients age 15-45, who were assigned female at birth, who are not currently pregnant, and who received contraceptive counseling as part of their recent visit.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, *describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

For the purposes of eligibility screening, patient age and sex are determined though patient report to their provider or clinic in the normal course of their care. As these are standard, readily available elements of patient data, clinics may rely on their own data to determine eligibility with regard to age and sex.

Receipt of contraceptive counseling is not a standard or readily available element of patient data. The current application presents data collected from patients responding to the PCCC immediately following their visit. Patients receiving contraceptive counseling during their visit are identified by providers and/or staff, following instructions provided by UCSF. Patient identification is then communicated to the team member responsible for distributing the PCCC survey to patients. Patients are identified through a standardized process that included pre-emptive staff review of schedules and visit types (e.g. flagging future family planning visits for survey distribution, as contraceptive counseling is likely to take place in such visits), and/or provider or staff identification based on the exam room conversation, depending on clinic protocols and flow. In the testing attachment we describe our assessment of the degree of ascertainment bias in this process.

As the PCCC is intended to measure the quality of counseling for those who did receive counseling, patients who did not receive counseling are not eligible to respond to the PCCC scale, regardless of whether counseling may have been appropriate during their visit. Whether or not people receive family planning care when appropriate is a distinct aspect of quality. This component of quality is partly captured by the existing NQF measure 2903, which assesses use of a most or moderately effective method. As all most or moderately effective methods require a prescription or a procedure from a provider, the score on this performance metric is influenced by the degree to which patients in need of family planning care receive these services. We acknowledge that future measures could be developed to more directly measure whether or not provision of contraceptive care is provided when appropriate.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Pregnant patients are excluded from the denominator, based on two reasons. First, contraceptive counseling in the context of pregnancy is distinct from that provided to non-pregnant individuals. Specifically, perinatal contraceptive counseling often includes multiple conversations touches over the course of prenatal care and immediate postpartum care. This is appropriate as women, when pregnant, are not immediately at risk of an undesired pregnancy, and therefore there is less time sensitivity to this counseling, and is also consistent with women's preferences for this care [1]. Given this difference in structure of counseling for pregnant women, the use of a visit-specific measure for contraceptive counseling is not appropriate.

Second, given distinct issues related to post-partum contraceptive use, including increased risk of blood clots, effect on lactation, and the health impact of birth spacing, counseling pregnant women about future contraceptive use has components distinct from that of non-pregnant women. For these conceptual reasons, the PCCC was designed for use with non-pregnant patients and has not been extensively tested with pregnant patients to determine whether it accurately captures their needs and desires for counseling.

References

[1] Yee LM, Farner KC, King E, Simon MA. What do women want? Experiences of low-income women with postpartum contraception and contraceptive counseling. Journal of Pregnancy and Child Health. 2015;2(5).

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

Staff and providers are instructed not to distribute the survey to patients whom have disclosed or discovered during the visit that they are pregnant. In addition, the survey asks patients if they are pregnant, and these responses are excluded from the calculation of the measure.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

We do not plan to stratify measure results in the current application. We plan to address stratification in maintenance applications for the measure, if applicable. We have collected data from all patients on their age, race, and ethnicity, and in the future we plan to address stratification by these categories. Please see testing attachment for our reasoning in delaying stratification to future maintenance applications.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Measure users should follow these steps in order to obtain measure results:

1) Identification and data collection

a) Providers and/or staff identify eligible, non-pregnant patients who have received contraceptive counseling, before they leave the clinic following their visit

b) A team member who is not the provider who gave counseling introduces and distributes the survey to the patient following their visit, before they leave the clinic

c) Patient completes the survey (self-administered via paper or electronically, e.g. on a tablet computer)

d) Electronic collection of patient responses for analysis, either through data entry of paper surveys or collation of responses to electronic survey

2) Data aggregation and measure calculation

a) Patients indicating they are pregnant have their responses excluded

b) Measure responses are summed as the total of all PCCC item values (maximum value of 20)

c) PCCC value sums are dichotomized as a maximum value of 20 (top-box score) versus any value less than 20

d) Dichotomized result variable is examined at the individual clinician/provider and facility level

e) Measure result is calculated as the percentage of patients responding with a top-box score, divided by the total number of patients who gave any response to the survey, on a provider or facility level

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

The PCCC is collected using consecutive sampling for identified providers and/or facilities. Based on our reliability results reported in the testing attachment, we recommend a minimum sample size of 30 responses on a provider-level, and 50 responses on a facility-level. While our results indicate a Spearman-Brown reliability >0.7 at both levels with 30 responses, we have chosen to be more conservative with facility-level recommendations due to both the lower value of the coefficient (0.78 compared to 0.84 at the provider-level), and the relatively minimal additional effort required at a facility-level to obtain 25 more surveys. No proxy responses are allowed.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

As described, patients receiving contraceptive counseling during their visit are identified by providers and/or staff, and patient identification is then communicated to the team member responsible for distributing the PCCC survey to patients. A team member who did not provide contraceptive counseling during the patient visit introduces and distributes the survey to patients, in order to mitigate the risks of social desirability bias and perceived threat to

privacy affecting patient responses, as could happen if the same individual who gave counseling were to distribute the survey. In both verbal and written communication to patients, the anonymous nature of the survey was emphasized, with assurance that answers to the survey would not be linked to the individual patient and would not impact care.

The survey is self-administered and returned before to the patient leaves the clinic. The response rate is calculated as the number of patients to whom a survey is administered who return a completed survey, divided by the number of patients approached to complete a survey. Consistent with the CAHPS recommendations for minimum response rates, we recommend a minimum response rate of 40% [1].

References

[1] Agency for Healthcare Research and Quality. CAHPS clinician & group survey and instruction. https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/survey3.0/fielding-the-survey-cg30-2033.pdf. Accessed 31 Jul 2019.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Instrument-Based Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

We used a brief patient survey including the PCCC in order to gather all data used in analyses. This survey is available in English and Spanish and is self-administered by patients (on a paper survey or electronically, e.g. on a tablet computer) immediately following the patient visit.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available in attached appendix at A.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician : Individual, Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

testing_attachment_PCCC_UCSF_1Aug2019_final.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

NATIONAL QUALITY FORUM — Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** N/A

Measure Title: The Patient-Centered Contraceptive Counseling (PCCC) scale

Date of Submission: August 1, 2019

Type of Measure:

Measure	Measure Continued
X Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing, (e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☐ Claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient survey responses	☒ other: Patient survey responses, audio recordings of clinic visits

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*)

N/A

1.3. What are the dates of the data used in testing? July 23, 2009 – June 20, 2019

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	Measure Tested at Level of:
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: NA	☐ other: NA

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1 below describes the facilities whose patients gave PCCC responses to be used in facility-level analyses. Tables 2 describes the types of individual and team-based clinicians/providers whose patients gave PCCC responses for provider-level analyses.

In recruiting facilities and clinicians/providers to this study, we wished to recognize the range of types of providers that engage in contraceptive counseling in order to ensure that our measure was

generalizable across settings employing different providers. For this reason, we included not only licensed clinicians, but also non-licensed providers, as facilities providing family planning care can employ either type of individual to provide contraceptive counseling services. We note that in cases where non-licensed providers provide counseling, licensed clinicians are always responsible for confirming the contraceptive method decision and ensuring a lack of contraindications. However, this process in itself does not constitute counseling, and therefore it is the counseling by the non-licensed provider that is the target of this measure in these settings.

In addition to the different roles of individuals providing counseling, some facilities use team-based models, in which two individuals (e.g., a nurse and a nurse practitioner) both provide elements of contraceptive counseling to a patient over the course of a visit (e.g. the nurse provides initial education and decision support, and the nurse practitioner then completes the counseling, including working with the patient to select the method). We therefore also included facilities with this model in our sample, and collected data in a manner that allowed us to take into account the team-based counseling model.

The PCCC was first collected from local facilities in the San Francisco Bay Area, beginning in 2009. In 2017, we began working with nine additional facilities from across the country to collect PCCC responses from their patients, in order to diversify the geographies from which we had data. At the same time, we shared the PCCC measure with the Oregon Health Authority (OHA) for use in facilities with which they work as the state agency that oversees state health-related programs. OHA included the PCCC in a brief patient experience survey with patients of participating facilities providing family planning services.

In all data collection efforts, survey questions were formatted similarly (see below under face validity in Section 2b1.2 for comparison of tablet-based and paper-based data collection). Due to different contexts and motivations for the data collection, there was some variation in the questions that participants were asked to answer prior to answering the questions that are a part of the PCCC, as would be expected in real world use. In assessing any potential impact these additional questions could have on our results, we saw no evidence of survey fatigue, as evidenced by no difference in lack of completion of the four items in the surveys in which there were additional questions compared to those without these questions.

In engaging with sites for the purpose of data collection, we wished to ensure we represented the range of contexts and processes in which contraceptive counseling is provided, in order to enhance generalizability. We therefore, as described, broadened the geographic area from which data was collected in the later phases of data collection. We also ensured inclusion of specialty family planning sites, including both Planned Parenthood and public health departments, as well as primary care sites, including Federally Qualified Community Health Centers.

Below are descriptive tables of the facilities (Table 1) and providers (Tables 2 and 3) included in the analysis).

Table 1. Facilities included in analysis

Facility	Location	Facility type	Patients served annually	Contraceptive counseling patients served weekly	Number of providers included in facility-level analysis	Number of providers included in provider-level analysis*	Rural/urban/suburban
1	Denver, CO	Primary Care Clinic (FQHC)	27,040	85	15	0	Urban
2	Milwaukee, WI	Family Planning	18,200	275	6	2	Urban
3	Carson City, NV	Public Health Department	5,200	30	11	7	Urban
4	Asheville, NC	Public Health Department	1,872	29	10	4	Urban
5	Page, AZ	Primary Care Clinic (FQHC)	22,360	18	8	2	Rural
6	Brooklyn, NY	Primary Care	5,200	50	5	0	Urban
7	Hyannis and Brockton, MA	Family Planning	7,800	90	5	3	Mixed

Facility	Location	Facility type	Patients served annually	Contraceptive counseling patients served weekly	Number of providers included in facility-level analysis	Number of providers included in provider-level analysis*	Rural/urban/suburban
8	Portland, OR	Family Planning	34,892	305	8	8	Mixed
9	Hempstead, NY	Family Planning	13,000	125	4	1	Suburban
10	Oregon	Family Planning	N/A**	36	N/A‡	0	Urban
11	Oregon	Public Health Department	N/A**	15	N/A‡	0	Urban
12	Oregon	Public Health Department	N/A**	21	N/A‡	0	Urban
13	Oregon	Family Planning	N/A**	25	N/A‡	0	Urban
14	Oregon	Public Health Department	N/A**	11	N/A‡	0	Rural
15	Oregon	Public Health Department	N/A**	10	N/A‡	0	Rural
16	Oregon	University Health Center	N/A**	20	N/A‡	0	Urban
17	Oregon	Primary Care Clinic (FQHC)	N/A**	27	N/A‡	0	Rural

Facility	Location	Facility type	Patients served annually	Contraceptive counseling patients served weekly	Number of providers included in facility-level analysis	Number of providers included in provider-level analysis*	Rural/urban/suburban
18	San Francisco, CA	Family Planning	18,200	65	7	4	Urban
19	San Francisco, CA	Community College-Based Clinic	13,000	13	6	0	Urban
20	San Francisco, CA	Public Health Department	17,107	57	5	1	Urban
21	San Francisco, CA	Obstetrics and Gynecology Clinic	109,200	N/A*	11	0	Urban
22	San Francisco, CA	Public Health Department	26,000	150	9	2	Urban

*Provider analysis only included those providers for whom at least 25 survey responses were collected.

**Annual data unavailable for these sites.

‡Survey responses for these sites not identifiable by provider.

Table 2. Roles of providers included in analysis

Provider role	n (%)
Medical doctor, nurse practitioner, certified nurse midwife, or physician assistant	**15 (44.1)**
Nurse	**3 (8.8)**
Non-licensed medical assistant, health educator, or other counselor	**10 (29.4)**
Two-person team consisting of one nurse practitioner and one nurse or medical assistant	**6 (17.7)**
Total	**34**

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Eligible patients were those who received contraceptive counseling during a visit with their health care provider. Contraceptive counseling was defined for care teams at participating facilities as follows:

> **What:** Any contraceptive education and/or discussion that may inform or influence a patient's choice of a contraceptive method. This includes deciding to continue with the same method or to use no method at all. This also includes discussion of just one or two methods; comprehensive counseling on all methods is not required.
>
> **Delivered by:** Any staff person who provides counseling as part of their job role, including advanced practice clinicians, clinical support staff, and non-clinicians (e.g., medical assistants, health educators).
>
> **Delivered to:** Any patient assigned female at birth who is not pregnant. This includes patients who identify as lesbian, gay, bisexual, or queer.
>
> **When:** During a patient visit of any type, except for prenatal visits or other visits with pregnant patients (for the rationale of exclusion of pregnant patients, please see section 2b2).

As the PCCC is intended to measure the quality of counseling for those who did receive counseling, patients who did not receive counseling were not considered eligible to respond to the PCCC scale, regardless of whether counseling may have been appropriate during their visit. Whether or not people receive this care when appropriate is a distinct aspect of quality. This component of quality is partly captured by the existing NQF measure 2903, which assesses use of a most or moderately effective method. As all most or moderately effective methods require a prescription or a procedure from a provider, the score on this performance metric is influenced by the degree to which patients in need of family planning care receive these services. We acknowledge that future measures could be

developed to more directly measure whether or not provision of contraceptive care is provided when appropriate.

Eligible patients were identified by staff and providers as having had received counseling according to the definition above. The identification of eligible patients was then communicated to the team member responsible for distributing the PCCC survey to patients. Patient identification was conducted using a standardized process that included pre-emptive staff review of schedules and visit types (e.g. flagging future family planning visits for survey distribution, as contraceptive counseling is likely to take place in such visits), and/or provider or staff identification based on the exam room conversation, depending on protocols and patient flow. In both verbal and written communication to patients, the anonymous nature of the survey was emphasized, with assurance that answers to the survey would not be linked to the individual patient and would not impact care. In order to mitigate the risks of social desirability bias in patient responses and perceived lack of privacy of responses among patients, the standardized process for survey distribution specified that the individual responsible for distributing the survey to patients was not the same team member who gave counseling.

Eligible patients were included in provider-level analyses if their provider saw 25 or more eligible patients who responded to the PCCC. Eligible patients were included in facility-level analyses if they visited a facility that had 25 or more eligible patients who responded to the PCCC. These thresholds were chosen based on simulations showing that 25 responses would result in adequately precise reliability estimates. To estimate the minimum panel sizes needed to obtain adequate precision in estimates of Spearman-Brown reliability (SBR) for our top-box scored binary outcome measure, we simulated correlated binary outcomes for range of patient panel sizes for samples of providers or clinics of varying sizes for a fixed value of the ICC based on preliminary data, then estimated the ICC from the simulated data using a mixed effects logistic model, and finally estimated SBR from the ICC and panel size, using the formula for SBR specified below. For each combination of sample and panel size, we repeated this procedure in 250 simulated datasets, then calculated the margin of simulation error (MSE) of the resulting SBR estimates, defined as 1.96 times the SD of the simulated SBR estimates. We considered an estimate MSE of +/- 7% to be acceptable.

Below are descriptive tables of patients included in the provider-level analyses (Table 4), facility level analyses (Table 5), and data element validity analyses (Table 6). Reasons for the different samples are described below.

Table 4. Demographics of patients included in provider-level analysis

Patient characteristics	n (%) for reliability analysis	n (%) for validity analysis
Age		
15-19	488 (19.7)	450 (19.8)
20-24	741 (29.9)	676 (29.8)
25-29	576 (23.3)	527 (23.2)
30-34	386 (15.6)	363 (16.0)
35-39	193 (7.8)	175 (7.7)

Patient characteristics	n (%) for reliability analysis	n (%) for validity analysis
40-45	90 (3.6)	79 (3.5)
Missing	3 (0.12)	1 (0.04)
Race		
African American/Black	222 (0.96)	198 (8.7)
American Indian or Alaskan Native	81 (3.3)	78 (3.4)
Asian	105 (4.2)	87 (3.8)
Pacific Islander	25 (0.1)	23 (1.0)
White	1335 (53.9)	1249 (55.0)
Mixed race/multiracial	156 (6.3)	146 (6.4)
Other*	147 (5.9)	134 (5.9)
Missing*	406 (16.4)	356 (15.7)
Ethnicity	NA	
Latina or Hispanic	747 (30.2)	660 (29.1)
Not Latina or Hispanic	1,717 (69.3)	1,599 (70.4)
Missing	13 (0.5)	12 (0.5)
Language		
English	2,275 (91.8)	2,093 (92.2)
Spanish	202 (8.2)	178 (7.8)
Birth control method chosen in visit	NA	NA
Pill	827 (33.4)	763 (33.6)
Patch	35 (1.4)	29 (1.3)
Ring	115 (4.6)	101 (4.4)
Injection	298 (12.0)	269 (11.8)
Intrauterine device (IUD)	428 (17.3)	398 (17.5)
Implant	303 (12.2)	280 (12.3)
Condoms	245 (9.9)	226 (10.0)
Withdrawal	36 (1.5)	36 (1.6)
Female sterilization	8 (0.3)	8 (0.4)
Male sterilization	9 (0.4)	9 (0.4)
Fertility awareness-based methods	9 (0.4)	8 (0.4)
Other	14 (0.6)	13 (0.5)

Patient characteristics	n (%) for reliability analysis	n (%) for validity analysis
None	71 (2.9)	65 (2.9)
Not sure	62 (2.5)	54 (2.4)
Missing	17 (0.7)	12 (0.5)
Total	**2,477**	**2,271**

*Most patients with race "Missing" or "Other" reported Latina/Hispanic ethnicity (96.3% of those with race "Missing," and 75.5% of those with race "Other").

Table 5. Demographics of patients included in facility-level analysis

Patient characteristics	n (%) for reliability analysis	n (%) validity analysis
Age		
15-19	717 (20.6)	601 (19.7)
20-24	1,066 (30.7)	931 (30.5)
25-29	796 (22.9)	712 (23.3)
30-34	485 (13.9)	449 (14.7)
35-39	273 (7.9)	245 (8.0)
40-45	132 (3.8)	116 (3.8)
Missing	9 (0.3)	1 (0.01)
Race		
African American/Black	350 (10.1)	343 (11.2)
American Indian or Alaskan Native	119 (3.4)	112 (3.7)
Asian	126 (3.6)	115 (3.8)
Pacific Islander	25 (0.7)	25 (0.8)
White	1,787 (51.4)	1,576 (51.6)
Mixed race/multiracial	244 (7.0)	175 (5.7)
Other*	185 (5.3)	181 (5.9)
Missing*	642 (18.5)	528 (17.3)
Ethnicity		
Latina or Hispanic	1,042 (30.0)	954 (31.2)
Not Latina or Hispanic	2,371 (68.2)	2,087 (68.3)
Missing	65 (1.9)	14 (0.5)
Language		
English	3,220 (92.6)	2,828 (92.6)
Spanish	255 (7.3)	227 (7.4)
Birth control method chosen in visit		
Pill	1,045 (30.1)	1,043 (34.1)
Patch	51 (1.5)	51 (1.7)
Ring	166 (4.8)	166 (5.4)
Injection	359 (10.3)	359 (11.8)
Intrauterine device (IUD)	524 (15.1)	521 (17.1)

Implant	**360 (10.4)**	**354 (11.6)**
Condoms	**294 (8.5)**	**291 (9.5)**
Withdrawal	**39 (1.1)**	**39 (1.3)**
Female sterilization	**8 (0.2)**	**8 (0.3)**
Male sterilization	**11 (0.3)**	**10 (0.3)**
Fertility awareness-based methods	**10 (0.3)**	**10 (0.3)**
Other	**27 (0.8)**	**27 (0.9)**
None	**88 (2.5)**	**87 (2.8)**
Not sure	**74 (2.1)**	**74 (2.4)**
Missing	**422 (12.1)**	**15 (0.5)**
Total	**3,478**	**3,055**

*Most patients with race "Missing" or "Other" reported Latina/Hispanic ethnicity (89.3% of those with race "Missing," and 77.6% of those with race "Other").

Table 6. Demographics of patients included in data element validity analysis

Patient characteristics	n (%)
Age	
15-19	41 (12.0)
20-24	115 (33.7)
25-29	88 (25.8)
30-34	40 (10.3)
35-39	37 (10.9)
40-53	20 (5.9)
Missing	0 (0.0)
Race	
African American/Black	98 (28.7)
Asian or Pacific Islander	0 (0.0)
White	243 (71.3)
Other	0 (0.0)
Ethnicity	
Latina or Hispanic	86 (25.2)
Not Latina or Hispanic	255 (74.8)

Birth control method chosen in visit	
Pill	119 (34.9)
Patch	7 (2.1)
Ring	47 (13.8)
Injection	33 (9.7)
Intrauterine device (IUD)	82 (24.1)
Implant	7 (2.1)
Condoms	23 (6.7)
Withdrawal	1 (0.3)
Female sterilization	2 (0.6)
Male sterilization	1 (0.3)
Fertility awareness-based methods	0 (0.0)
Other	17 (5.1)
None	0 (0.0)
Not sure	1 (0.3)
Missing	0 (0.0)
Total	**341**

If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The dataset used for provider-level reliability analyses is smaller than that used for facility-level analyses, as the facility-level analyses includes all provider-level data, and provider-level data only includes those providers for whom we had an adequate number of patient responses. However, within each level of reliability analyses, the data used is consistent throughout.

Performance measure validity was tested using the sub-set of the reliability datasets for which we had collected two pre-specified patient-reported validity items immediately following completion of the PCCC items. For provider-level analyses this included 2271 patients and 34 providers from 13 facilities. For facility-level analyses this included 3055 patients and 116 providers from 15 facilities.

Critical data element validity was tested using a dataset of 341 patients and 38 providers who had had their clinic visits audio recorded.

1.8. What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not

collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

We collected data on race/ethnicity for all individual patients.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Critical data elements: Reliability testing for the critical data elements was examined with Cronbach's alpha. Cronbach's alpha is a measure of internal consistency that examines how closely related a set of items (critical data elements) are as a group or the extent to which the data elements measure the same concept or construct.

The formula for Cronbach's alpha is $\alpha = N\bar{c}/(\bar{v} + (N-1)\bar{c})$, where N is the number of items, $\bar{c}$ is the average between-item covariance, and v is the average within-item variance.

Performance measure score: As described in the Intent to Submit form, for both conceptual and analytic reasons, our calculation of the performance score used a dichotomous scoring system, in which all surveys in which the highest rating was given for all four questions are considered a positive score, whereas any survey in which a less than optimal rating on any of the four questions is considered a negative score. To assess reliability of this measure, we adhered to the NQF recommendations in the commissioned paper entitled "Patient-Reported Outcomes in Performance Measurement," in which signal-to-noise ratio (SNR) is recommended as a measure of reliability,[1] with signal defined as the variance in a performance measure due to systematic differences across units, and noise as the residual variance due to random error within units. We have focused on the equivalent Spearman-Brown (S-B) measure of reliability,[2-4] which is a function of the intraclass correlation (ICC), defined as the ratio of between-unit variance to the sum of between- and within-unit variances, and the prospective panel size to be used in evaluating facility and/or provider performance. Specifically, with a prospective panel size of n and $ICC = V_b/(V_b + V_w)$, where V_b and V_w denote the between- and within-unit variances, we have

$$
\begin{aligned}
\text{Reliability}_{S-B} &= nICC/(1+(n-1)ICC) \\
&= nV_b/(V_b+V_w)/(1+(n-1)V_b/(V_b+V_w)) \\
&= nV_b/(nV_b+V_w) \\
&= V_b/(V_b+V_w/n) \\
&= \text{SNR}
\end{aligned}
$$

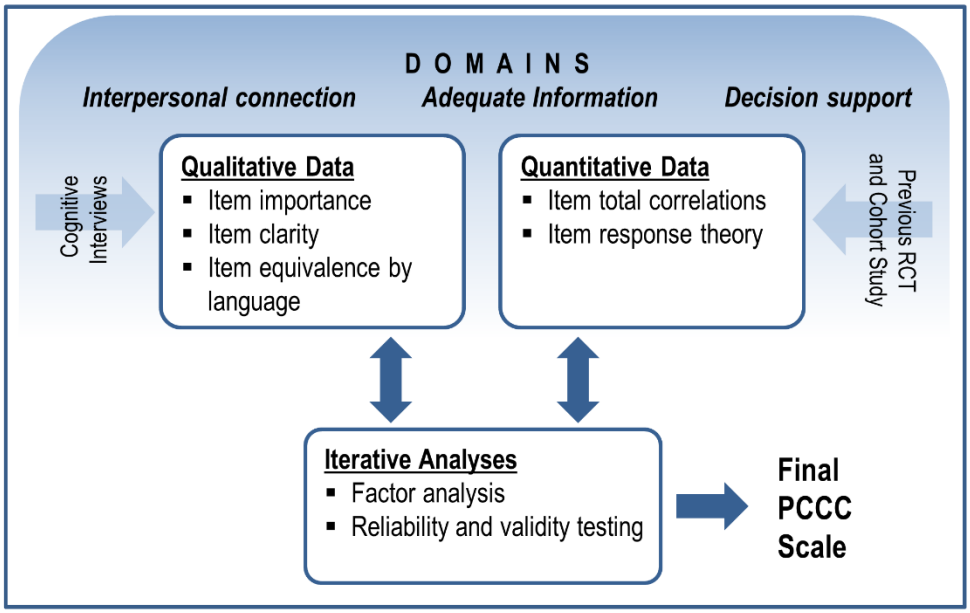
With a binary performance measure, and respondents nested within providers, who are in turn nested within facilities, the ICC can be estimated using a normal-logistic model with nested random effects for facility and provider. Specifically, our analysis used the Stata **`melogit`** and **`estat icc`** commands to estimate the ICC with 95% confidence intervals. Confidence bounds for S-B reliability were estimated by plugging the confidence limits for the ICC provided by **`estat icc`** into the formula for the Spearman-Brown reliability. Our analysis was implemented using Stata Version 16.0 (StataCorp LLC, College Station, TX 77845).

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

The PCCC scale that we are submitting for endorsement is a shortened version of a scale (the Interpersonal Quality of Care scale [IQFP]) that our team originally developed as a patient-reported outcome measure in the context of contraceptive research. The IQFP was designed to capture the three domains of quality contraceptive counseling found to be important to patients in qualitative research, including interpersonal connection between provider and patient, adequate information, and decision support.[5] The final 11 items included in the IQFP were selected using factor analysis from 17 initial items, in addition to be a reliable measure with content, construct, convergent, predictive and discriminant validity.[6]

In order to shorten the IQFP for use as a performance measure, we used a process in which qualitative data addressing item importance and interpretability, as well as equivalence between English and Spanish, was triangulated with quantitative data addressing reliability and validity (see Figure 1). This process led us to a 4-item scale retaining the validity and reliability of the 11-item IQFP. Reliability and validity testing of the 4-item patient-reported outcome measure conducted with the original scale during the development and item reduction process will be described in more detail in documents to be submitted later in the endorsement application process.

Figure 1. Qualitative and quantitative data triangulation for item reduction of the initial IQFP to the PCCC



Using the datasets described in section 1.7, we calculated the Cronbach's alpha for the PCCC. Using data available for testing at the provider level, the PCCC measure has a Cronbach's alpha of 0.94 (Item-total correlations ranged from 0.80 - 0.89). Using data available for testing at the facility level, the PCCC measure has a Cronbach's alpha of 0.93 (Item-total correlations ranged from 0.78 - 0.87). Inspection of Cronbach's alpha if any single item was deleted yielded no improvement in Cronbach's alpha for available data at the provider or the facility level.

Table 7 displays the provider- and facility-level estimates of Spearman-Brown reliabilities for a range of panel sizes. Our ICCs of 0.10-0.15 result in adequate reliability with moderate panel sizes of 20-50. In particular, the estimated reliability of samples of 50 surveys at a facility level is 0.85, while this value is 0.84 at a provider-level with a panel size of 30.

Table 7: Spearman-Brown Reliabilities of PCCC, by Prospective Panel Size

Panel Size	**Provider-Level Reliabilities (95% CI)**	**Facility-Level Reliabilities (95% CI)**
25	0.81 (0.63, 0.92)	0.74 (0.59, 0.86)
30	0.84 (0.67, 0.93)	0.78 (0.63, 0.88)
50	0.90 (0.77, 0.96)	0.85 (0.74, 0.92)
100	0.95 (0.87, 0.98)	0.92 (0.85, 0.96)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

First, regarding Cronbach's alpha, a value for Cronbach's alpha of 0.70 or above is viewed as acceptable and a value of 0.90 or above is excellent and indicating a strong internal consistency of the measure.[7,8] Our reported values of 0.94 and 0.93 for the available data at the provider level and facility level respectively are in the excellent category, especially in light of the small number of items (4 items) providing assurance the high Cronbach alpha level is not inflated by instrument length.

Our reliabilities of >0.7 for panel sizes of 25-50 at both provider- and facility-level are consistent with recommendations for reliability estimates for performance measurement.[4,9] These values also compare favorably with reliability estimates for the CG-CAHPS surveys, including the communication composite score, which has been reported to have a reliability of 0.62-0.81.[10] Based on these results, we recommend a minimum panel size of 30 for provider-level assessment. We have chosen to be more conservative with facility-level recommendations due to both the lower value of the coefficient (0.78 compared to 0.84 at the provider-level with a panel size of 30), and the relatively minimal additional effort required at a facility-level to obtain 20 more surveys.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*). **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Critical data elements: We evaluated PCCC critical data element validity by assessing the association between each of the four critical data elements (individual PCCC items) with specific clinician communication practices consistent with patient-centered care and assessed from audio recordings of visits using measures derived from the Four Habits Coding Scheme (4HCS). The 4HCS is a validated observational approach to assessing patient-centered health communication including systematic coding,[11] which we modified slightly for the family planning context in collaboration with the original developer of the measure.[12] Specific items or components of the 4HCS were selected to test for associations with each of the four PCCC individual items on the basis of conceptual decisions and content. Linear mixed models (LMMs) assessed the association between individual PCCC items used in a continuous (1-5 response scale) and the specified 4HCS components (aggregated across specified items) and adjusting for clustering by provider (random effect).

Performance measure score:

Empirical validity testing: Convergent validity at the performance measure level was assessed by examining associations between the PCCC and two patient-reported measures using provider-level averages: (1) satisfaction with provider help with the choice of a birth control method and (2) satisfaction with the method choice. These two single items self-report measures were collected using the same modality as the PCCC scores. The choice of these measures of validity was based on the fact that measures of satisfaction often correlate with measures of patient-centered processes of care but are considered distinct. We conceptualized the PCCC as being more specific than measures of satisfaction, as satisfaction measures tend to be informed by expectation disconfirmation theory (i.e., the extent to which an experience exceeded or fell below expectations[13,14]) and have additional limitations of lack of differentiation and lack of specificity of measured behaviors.[15] In contrast, the items in the PCCC assess the extent to which the patient experienced or perceived specific types of communication and exchanges consistent with patient-centered care. To measure satisfaction with provider help with birth control choice, women were asked to rate their "How satisfied are you in how your healthcare provider helped you to choose what birth control method to use" on a 7-point Likert scale from excellent to poor. Satisfaction with the method choice was assessed using the question "How satisfied are you with your choice of birth method at this visit?", using a 7-point Likert scale. In line with our previous PROM-level work and to allow for aggregating data across alternative 5-point Likert versions of the items, patients in the highest or most positive rating category were compared to all others. Associations between the PCCC and the convergent validity measures were assessed using linear mixed models (LMMs) using provider-level and facility-level averages to examine validity at the provider- and facility-level respectively, with random facility effects included for provider-level averages.

Systematic assessment of face validity: We assessed face validity of the performance measure score with facility administrators, providers of contraceptive counseling, and patients. We assessed face validity with administrators and providers by conducting two Modified Delphi Processes via e-mail (one with a group of 14 administrators, and one with a group of 19 providers) with participants from facilities across the country. Each Modified Delphi Process used two rounds of both close-ended and open-ended questions to collect feedback from each group on the performance measure, with the second round of questions reflecting feedback from the first round in order to move towards consensus on face validity. Each Process asked participants to reflect on the acceptability of items, whether they thought a dichotomized top-box score of the item responses would accurately reflect their performance, and the applicability and utility of a performance score to their work. Each Modified Delphi Process resulted in consensus that the performance measure score was useful in differentiating high quality and lower quality care, and that the performance measure score would be of use to the work of administrators and providers. Among providers, 90% indicated that they would be likely to consider a provider receiving a higher score on this measure to be providing better care (giving a response of at least 7 on a scale from 1 to 9, from very unlikely to very likely). Ninety-two percent of administrators gave a response of at least 7 on the same item. With regard to usefulness, 88% of providers and 93% of administrators agreed, based on a response of at least 7 on a scale of 1 to 9, that reporting the percentage of responses that were top-box scores would be understandable as an indicator of performance and meaningful for quality improvement.

We conducted interviews and focus groups with 43 patients from across the country to assess the face validity of the measure with this group. In order to obtain diverse representation in this qualitative sample, we recruited focus group and interview participants who had recently experienced contraceptive counseling with the assistance of health care facilities serving diverse patient populations in three states (Texas, North Carolina, and California). We asked patients about the acceptability of items and the utility of a dichotomized top-box score of item responses for their decision-making about their health care. We also

assessed any differences in patient preference and response between paper and electronic versions of the survey. Interviews and focus groups revealed that patients found the items and performance measure score were acceptable and useful to them. Eighty-eight percent of patients reported that a facility or provider having a higher score on the performance measure would make them more likely to choose that facility or provider for their care, as opposed to having no effect on their decision-making about their care. Patients responded to the measure in an equivalent fashion in both paper and electronic versions. The equivalence in patient response to paper and electronic versions of the survey was further supported in data collection for reliability and validity testing. One facility in this testing used both paper and electronic surveys with different patients (implementing first electronic and then paper surveys in sequence), and very similar percentages of patients gave the top-box score by each modality in this facility (87.5% on paper and 86.1% electronically [difference not significant]).

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Critical data elements: Analyses were based on the available sample in which the 4HCS observations and coding data were collected (patient n=341). The sample (described in table 6), represents four publically funded clinics, with a sample of women that was diverse in terms of race and age. Per the 4HCS coding system, 4HCS components are interpreted in the direction of lower values indicative of highly effective use of a code, whereas higher values are indicative of lower effective use. As shown in Table 9, each critical data element from the PCC was significantly associated with the specific 4HCS component selected on the basis of conceptual match, with the negative betas signifying the association between higher PCCC scores and lower (meaning more effective use) of the 4HCS component. Betas presented in Table 9 represent a unit change in the PCCC item score for each unit change in the 4HCS score. All results remained significant and a virtually identical pattern of findings resulted when testing the PCCC items dichotomized (5 vs. else) and when not adjusting for provider.

Table 9: Validity testing of critical data elements (n=341)

PCCC item	4HCS component	B (SE)	95% CI	p value
Respecting me as a person	Invest in the beginning	-1.13 (0.32)	-1.76 – -0.51	<0.001
	Demonstrate empathy	-1.22 (0.18)	-1.57 – -0.87	<0.001
Letting me say	Elicit the patient perspective	-0.55 (0.12)	-0.80 – -0.30	<0.001
Preferences	Elicit the patient perspective	-0.59 (0.13)	-0.86 – -0.32	<0.01
Giving information	Information provision	-0.22 (0.84)	-0.38 – -0.05	0.01

Note: The following individual codes are included in each of the following 4HCS components (items scored from 1 to 5 and summed to calculate component scores): Invest in the beginning (shows familiarity, greets patient warmly, makes small talk, uses primarily open ended questions, encourages expansion of medical concern, elicits full range of concerns, possible range of scores 6 - 30); Demonstrate empathy (encourages appropriate expression of emotion, shows empathy for patient experience/feelings, helps to identify/label feelings, possible range of scores 3 - 15); Elicit the patient perspective (elicits patient's experiences around birth control, elicits patient's preferences around birth control, shows interest in impact on patient's life of birth control use, possible range of scores 3 - 15); and Information provision was created from two items

specifically for the current analysis to conceptually align with providing information (provided personalized information about options, explains rationale about plan, possible range of scores 2 - 10).

Performance Measure: Analyses were based on the provider and facility datasets defined earlier. Results indicated the PCCC and two self-reported validity items are strongly associated at both the provider and facility level, comparing percentages of top-box scores on the PCCC with percentages of top-box scores on the single-item validity measures of satisfaction. In zero-order correlations, at the provider level aggregated percentage of high PCCC scores are positively associated with method choice satisfaction (r=0.82) and satisfaction with provider help with birth control choice (r=0.88) both p<0.001. Likewise, at the facility-level, aggregated PCCC scores are associated with r=0.76 with method choice satisfaction and r=0.82 for satisfaction with provider help (n=15 facilities; both p<0.001). Linear mixed model results, adjusting for facility in provider level analyses, mirror these findings and are summarized in Table 10 below, with higher rates of high PCCC scores for a provider or facility associated with higher rates of satisfaction on both measures.

Table 10: Validity testing compared to single item measures, on both provider- and facility-level

Level of analysis	Validity item	B (SE)	95% CI	p value
Provider	Satisfaction with method choice	1.01 (0.18)	0.66 – 1.35	<0.001
Provider	Satisfaction with provider help with method choice	0.89 (0.58)	0.78 – 1.00	<0.001
Facility	Satisfaction with method selected	1.53 (0.36)	0.75 – 2.30	0.001
Facility	Satisfaction with provider help with method choice	1.02 (0.20)	0.59 – 1.44	<0.001

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Our results for validity of critical data element validity demonstrate that our survey items are associated with the patient-centeredness of contraceptive counseling discussions measured through observation of visits using audio recordings. This indicates that patients' subjective experience of patient-centeredness reflects the observed quality of counseling. Our empirical validity testing of the performance score documents that our measure is correlated with other conceptually related measures of quality. Our rigorous process of face validity testing with providers, administrators, and patients provides further supports that the PCCC measure provides an accurate reflection of patient-centered contraceptive counseling and can be used to distinguish good from poor quality contraceptive counseling.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We excluded pregnant people who responded to the survey from our analyses, based on two reasons. First, contraceptive counseling in the context of pregnancy is distinct from that provided to non-pregnant individuals. Specifically, perinatal contraceptive counseling often includes multiple conversations touches over the course of prenatal care and immediate postpartum care. This is appropriate as the woman, when pregnant, is not immediately at risk of an undesired pregnancy, and therefore there is less time sensitivity to this counseling, and is also consistent with women's preferences for this care.[16] Given this difference in structure of counseling for pregnant women, the use of a visit-specific measure for contraceptive counseling is not appropriate.

Second, given distinct issues related to post-partum contraceptive use, including increased risk of blood clots, effect on lactation, and the health impact of birth spacing, counseling pregnant women about future contraceptive use has distinct components than that of non-pregnant women. In future work, we are interested in developing and validating an equivalent measure for pregnant women that captures their experience of contraceptive counseling across their perinatal care.

A total of 45 pregnant people otherwise eligible for the PCCC responded to the measure (which equates to 1.8% of those responding for the provider level of analysis, and 1.3% of those responding for the facility level of analysis). We ran sensitivity analyses including pregnant people in analyses of performance measure descriptive statistics.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

We observed no differences in results in aforementioned sensitivity analyses.

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

In addition to being consistent with the manner in which this measure was developed and tested, the exclusion of pregnant people has essentially no impact on measure results.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with __ risk factors**

☐ **Stratification by ___ risk categories**

☐ **Other, **

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

N/A

2b3.2. If an outcome or resource use component measure is *not risk adjusted or stratified,* provide *rationale and analyses* to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

We do not believe that risk adjustment is justified. While it is possible that different demographic groups may report different results on the PCCC measure, this would represent true differences in patient-centeredness due to the manner in which the questions are framed and the fact the concepts of respect and attention to preferences and adequate provision of information are generally desirable.

With respect to the question of stratification, we do note that studies have suggested that women of color receive poorer quality contraceptive counseling than their White counterparts.[17] These disparities are rooted in the long history in the United States of coercion on the part of the reproductive health care system towards women of color, including forced sterilization and pressure to use long-acting contraceptive methods.[18,19] While this suggests that stratification by race/ethnicity may be desirable in order to assess differences in care, the PCCC has not yet been evaluated to assess these differences in an accurate and nuanced way. We acknowledge the existence of these disparities and intend to use the PCCC to examine them more closely in the future. In the current application, we wish to demonstrate the validity and reliability of the measure overall before taking a focused approach to examining disparities in quality of care.

2b3.3a. Describe the conceptual/clinical <u>and</u> statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

N/A

Also discuss any "ordering" of risk factor inclusion; for example, are social risk factors added after all clinical factors?

N/A

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

- ☐ **Published literature**
- ☐ **Internal data analysis**
- ☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

N/A

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

N/A

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow*

statistic): **2b3.8. Statistical Risk Model Calibration – Risk decile plots or**

calibration curves: **2b3.9. Results of Risk Stratification Analysis**:

N/A

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified

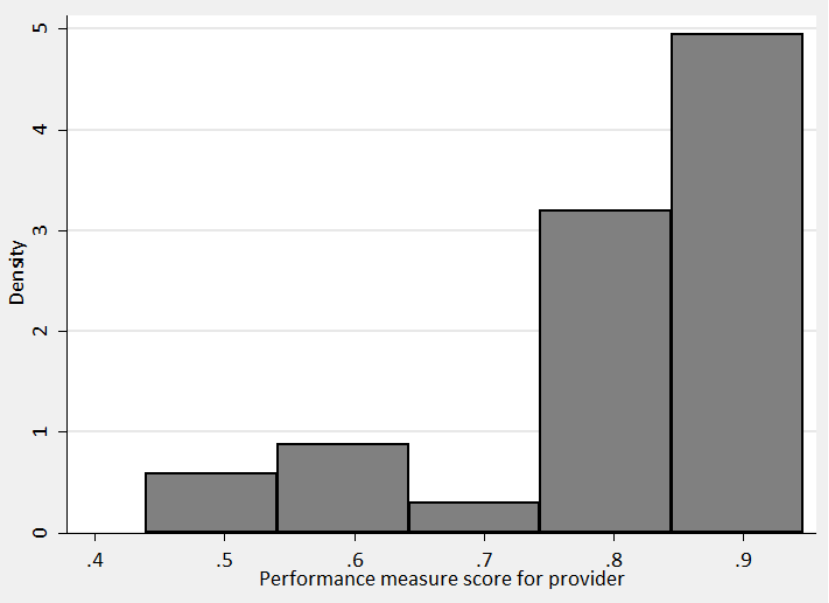
(*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

We examined descriptive statistics (including mean, standard deviation, range, and percentile scores) to understand the distribution of provider-level and facility-level scores.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

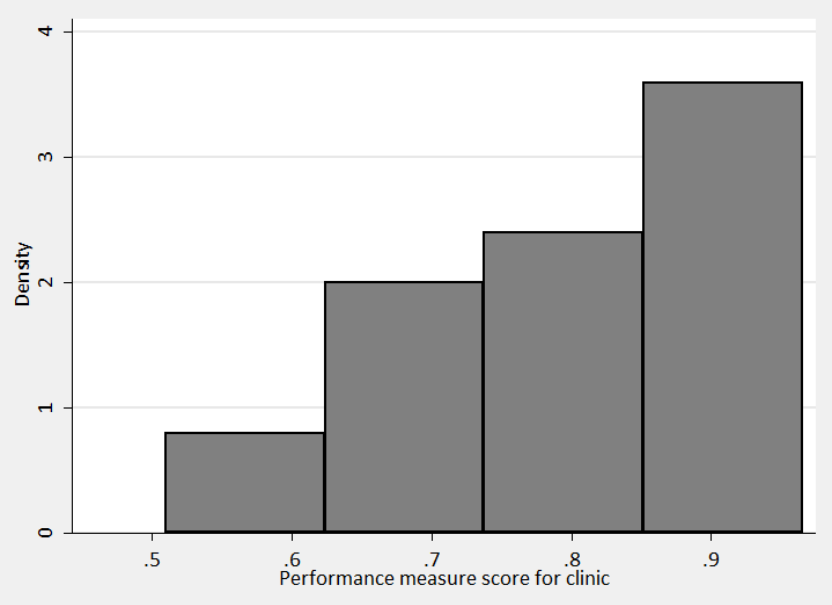
- Provider-level analysis (n=34)
  - Mean performance score: 0.81
  - Standard deviation: 0.12
  - Range: 0.44-0.95
  - Percentiles
    - 25^{th}: 0.79
    - 50^{th}: 0.85
    - 75^{th}: 0.90

Figure 2. Histogram of provider performance measure scores



- Facility-level analysis (n=22)
  - Mean: 0.79
  - Standard deviation: 0.12
  - Range: 0.51-0.97
  - Percentiles
    - 25th: 0.70
    - 50th: 0.83
    - 75th: 0.88

Figure 3. Histogram of facility performance measure scores



2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

We found that providers and facilities had a median score of 86% and 83% respectively, with 25th/75th percentile values of 79%/90% and 70%/88% top-box responses and ranges extending towards 50%. This indicates that while the performance measure scores are right-skewed in a manner similar to many patient satisfaction and experience measures, there is variability in scores indicating the opportunity for improvement among low performers. Of note, our distribution is wider than that of available statistics on the CG-CAHPS communication composite score, in which the median score at a clinic level is 88%, with a 25th/75th percentile of 84%/91%.[20]

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic

missing data (or differences between responders and nonresponders) and how the specified handling of
missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

In order to assess the occurrence of missing data due to eligible patients missed in patient identification, the UCSF project team made site visits to a selection of four participating sites in Fall of 2018. During site visits, UCSF team members conducted a total of 28 interviews with site administrators, providers, and staff. Providers were specifically asked to describe their process of identifying patients to complete the PRO-PM, in order to understand whether any bias existed in their patient selection. The UCSF team also conducted chart reviews of 300 charts at each of the four sites visited. The UCSF team reviewed charts of a total of 1,200 patient charts on site visits. Data was recorded on patient age, provider, type of visit, and whether the patient received counseling or not. Sensitivity and specificity were calculated by site to understand the accuracy of patient identification. Age and visit types were compared between true-positive and false-negative cases of patients who received counseling using chi-squared tests to understand if patients were missed systematically.

To evaluate overall differences between those who completed the survey and those served by the sites at which data collection occurred, we again used a chi-squared test to compare the race/ethnicity of survey respondents to those of all patients served by each facility, as reported by the facilities at the beginning of data collection.

We also examined both item non-response (i.e., incomplete responses to the four-item PCCC) and survey non-response (i.e., patients who were approached to take the PCCC survey but declined to do so). While item non-response rates could be calculated for all facilities and providers, survey non-response rates were only available for the nine facilities in our nationwide sample. At those sites, sequential IDs for both paper and electronic surveys were assigned to each patient, including those who declined the survey, allowing us to tabulate survey non-response. We assessed whether there were statistically significant differences in response rates between providers and between facilities using chi-squared tests. In order to test what effect biased non-response would have on performance scores, we imputed the missing data by provider and facility under the conservative assumptions that all individuals with missing data were either 25% more or less likely to give a top-box score than were respondents.

To assess sensitivity of our complete-case estimates for reliability to both item and survey non-response, we imputed summary PCCC scores for an additional 15% of patients in the panel for each provider in the complete dataset, based on the overall combined item-level and survey-level non-response rate in the nationwide sample. Specifically, we imputed 0% to 100% of the missing responses as top-box scores, with the remainder as non-top box scores, to check sensitivity of our reliability testing (described in section 2a2.2) to missingness driven by the care experience.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

In our chart review to assess the impact of eligible patients not receiving the survey, we found that in each site, about 50% of patients who received counseling did not receive a survey (false-negative). When true-positive and false-negative cases were compared, there were no systematic differences in visit type or patient age. This aligned with our findings from the interviews with staff, administrators, and providers, in which these individuals reported that failing to distribute surveys to patients was usually a result of clinic factors (e.g. related to clinic flow) and not individual patient factors (e.g., patient demographics, length of visit, or relationship with patient). In our analysis comparing the race/ethnicity of survey respondents and overall patient populations at facilities, we observed no statistically significant differences by facility between the races and ethnicities of overall patient populations and survey respondents.

With respect to item-level missing data, our overall rate was 1.7%, with a range between 0 and 10% for both providers and facilities. Assessing survey-level non-response data from the nine clinics participating in our nationwide test, we observed an overall rate of missingness of 13%, varying from 0 to 50% by provider and 0 to 32% by facility; these differences were not statistically significant. In our sensitivity analysis in which we imputed missing data as being 25% more or less likely to be top-box scores for providers and facilities in the nine sites for which we had survey-level non-response data, we found that there was low potential for biased non-response to substantially impact the assessment of the quality of care. Even under the conservative assumption in which all non-responders were eligible to complete the survey and were 25% more likely to give a top-box score, our imputed performance score was only 0-4% higher than our observed data for both providers and facilities. If we imputed responses being 25% less likely to give a positive response, for provider scores were 0-8% lower, and for facilities 0-7% lower.

In our sensitivity analyses using imputed data in the entire sample to assess the effect on reliability, our estimated facility-level Spearman-Brown reliability with a panel size of 50 was robustly >0.7 across the range of imputed top-box score rates. With a provider-level panel size of 30, the reliability estimate remained >0.7 with top-box scores imputed for at least 50% of non-respondents, and was only mildly sensitive to lower imputed top-box score rates, with an estimated reliability of 0.67 when top-box scores were imputed for only 10% of non-responders.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

It is our assessment that a 50% rate of distributing surveys to potentially eligible patients is realistic and acceptable for a typical health center setting, as long as patients are missed due to facility and contextual factors and not individual patient-level factors. Both our quantitative and qualitative data indicates that bias in survey distribution is likely to be minimal, with patients not being missed systematically by factors such as experience of counseling, demographics, or appointment type.

Our low item-level non-response and survey-level non-responses rates, especially compared to other patient experience surveys such as CG-CAPHS,[21] decreases the likelihood that our complete-case estimates of reliability are meaningfully biased. While there is some variation across providers and site in survey level non-response, we did not find this to be statistically significant. Our imputation of missing data in the nine sites for which we had provider- and facility-level information under assumptions of biased non-response indicates that missing data is not likely to result in meaningful under or over-estimates of provider performance.

Our reliability estimates also appear fairly insensitive to a range of assumptions about informatively missing data. Specifically, facility-level reliability with panel sizes of 50 never dropped below 0.7 in the imputed data, and provider-level reliability with panel sizes of 30 only dropped below 0.7 if non-respondents were much more likely to have had a poor experience than those who did respond, and even then only dropped to 0.67 under the fairly extreme assumption that only 10% or fewer non-respondents would have given their providers top-box scores.

References:

1. Deutsch A, Smith L, Gage B, Kelleher C, Garfinkel D. Patient-reported outcomes in performance measurement. Commissioned paper prepared for the National Quality Forum. 2012; Washington, DC.
2. Spearman C. Correlation calculated from faulty data. *British Journal of Psychology.* 1910;3:271-295.
3. Brown W. Some experimental results in the correlation of mental abilities. *British Journal of Psychology.* 1910;3:296-322.
4. Eijkenaar F. Profiling individual physicians using administrative data from a single insurer. *Medical Care.* 2013;51:731-739.
5. Dehlendorf C, Levy K, Kelley A, Grumbach K, Steinauer J. Women's preferences for contraceptive counseling and decision making. 2013;88(2):250-256.
6. Dehlendorf C, Henderson JT, Vittinghoff E, Steinauer J, Hessler DJC. Development of a patient-reported measure of the interpersonal quality of family planning care. 2018;97(1):34-40.
7. Cortina JM. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology.* 1993;78(1):98.
8. Nunnally JC. *Psychometric theory.* 2nd ed. New York: McGraw-Hill; 1978. Chapter: Assessment of reliability; p.245-24.
9. Adams JL, Mehrotra A, Thomas JW, McGlynn EA. Physician cost profiling--reliability and risk of misclassification. *New England Journal of Medicine.* 2010;362(11):1014-1021.
10. Dyer N, Sorra JS, Smith SA, Cleary PD, Hays RD. Psychometric properties of the consumer assessment of healthcare providers and systems (CAHPS(R)) clinician and group adult visit survey. *Medical Care.* 2012;50 Suppl:S28-34.
11. Krupat E, Frankel R, Stein T, Irish J. The four habits coding scheme: validation of an instrument to assess clinicians' communication behavior. *Patient Education & Counseling.* 2006;62(1):38-45.
12. Dehlendorf C, Henderson JT, Vittinghoff E, et al. Association of the quality of interpersonal care during family planning counseling with contraceptive use. *American Journal of Obstetrics and Gynecology.* 2016;215(1):78 e71-79.
13. Kupfer JM, Bond EU. Patient satisfaction and patient-centered care necessary but not equal. *Journal of the American Medical Association.* 2012;308(2):139-140.
14. Williams S, Weinman J, Dale J, Newman S. Patient expectations: what do primary care patients want from the GP and how far does meeting expectations affect patient satisfaction? *Family Practice.* 1995;12(2):193-201.

15. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N. Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings. Paper prepared of the National Quality Forum. 2012; Washington, DC.
16. Yee LM, Farner KC, King E, Simon MA. What do women want? Experiences of low-income women with postpartum contraception and contraceptive counseling. *Journal of Pregnancy and Child Health.* 2015;2(5).
17. Becker D, Tsui AO. Reproductive health service preferences and perceptions of quality among low-income women: racial, ethnic and language group differences. *Perspectives on Sexual and Reproductive Health.* 2008;40(4):202-211.
18. Stern AM. Sterilized in the name of public health: race, immigration, and reproductive control in modern California. *American Journal of Public Health.* 2005;95(7):1128-1138.
19. Roberts D. *Killing the black body: race, reproduction, and the meaning of liberty.* New York, NY; Pantheon Books, 1997.
20. Agency for Healthcare Research and Quality. CAHPS aggregated data. https://cahpsdatabase.ahrq.gov/CAHPSIDB/CG/Percentile.aspx. Accessed June 30, 2019.
21. Cleary P, Elliott M. Sorting fact from fiction: the value of patient experience measurement. Webinar hosted by the Agency for Healthcare Research and Quality. 2016; Arlington, VA.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Patient/family reported information (may be electronic or paper)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

The PCCC is a PRO-PM; both its numerator and denominator are calculated based on patient response to the PCCC survey, and thus cannot be captured from existing electronic sources. Patients' PCCC responses may be captured electronically using patient-facing interfaces such as kiosks or tablet computers in the clinical setting. Responses may also be collected using paper surveys with patients. Facilities may select a modality of data collection (electronic vs. paper) depending on capacity, resources, and workflows.

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. <u>Required for maintenance of endorsement.</u> Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

<u>IF instrument-based,</u> consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

We worked with facilities across the United States to implement the PCCC for the purposes of validity and reliability testing. In doing so, we were able to benefit from and build upon our partner organizations' existing strengths in implementing patient surveys, as well as work with partners to address challenges and find solutions in order to complete data collection. Our collaboration with partners allowed for iterative development of best practices for survey administration, which we have compiled in an implementation manual to help guide future use of the PCCC.

Partner organizations' iterative input related to project feasibility and existing survey practices guided the development of a standard but flexible workflow for survey implementation, which all participating sites used in validity and reliability testing. The resulting process consists of staff and providers identifying patients eligible for the survey at the time of their visit and distributing the survey to the patient in the health care facility at the end of the appointment, before they left the site. In piloting of survey distribution strategies prior to official data collection, this workflow was more feasible for most partners and produced higher response rates compared with EHR-based identification and post-visit survey distribution, which our team had previously considered as an implementation strategy.

In the Fall of 2018, we selected four partner organizations with which to conduct information-gathering activities to better understand challenges to PCCC implementation. We conducted interviews with administrators, providers, and staff of these organizations about challenges and barriers to implementation and participant opinions of the PCCC testing project. The biggest challenge to implementation using the standardized workflow was the inconsistency of staff and providers remembering to indicate the appropriate patients to take the survey, in the midst of the busy clinical environment. Based on these results, we produced messaging and materials to aid clinical leadership in reminding providers and staff about the project, e.g. more frequent and detailed reports of project progress to share at staff meetings. Respondent opinions of the PCCC and pilot were positive, indicating little risk that negative provider/staff attitudes toward the PCCC would influence the success of data collection.

With respect to the patients completing the survey themselves, as described in our testing attachment, we conducted face validity testing with patients to explore their feelings about completion of this survey and worked to minimize burden through such mechanisms as attention to survey formatting and reduction of the number of items to 4 (from the original 11). Our high response rate (87%) in real world testing of the PCCC indicates that this survey is feasible and acceptable for patients to complete.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Quality Improvement (Internal to the specific organization) Oregon Health Authority URL not currently available

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

The state-sponsored Oregon Health Authority (OHA) initiated use of the PCCC on its family planning patient experience survey in 2018. A survey including the PCCC was implemented for several weeks in Spring 2018 in 41 clinical sites across Oregon, including public health departments and non-profit clinical facilities, in order for OHA and facilities to gain a snapshot of patient experience and inform quality improvement. Results were calculated at the facility level.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

PCCC results are not currently publicly reported or used in an accountability application. This is because the PCCC measure has not yet been widely disseminated or publicized, and thus has not yet been widely used. We have been in communication with relevant organizations and agencies, such as NCQA, the Bureau of Primary Health Care, and OPA throughout our process of measure development and validation. Following measure endorsement, we plan to engage further with these groups to facilitate dissemination and use overall, and specifically with respect to public reporting and Patient-Centered Medical Home (PCMH) certification. There are no policies or entities that currently stand to restrict access to results or impede implementation.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program,*

purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)

As part of our planned project with FQHCs, we intend to work toward incorporation of the PCCC into requirements for PCMH certification within three years. We also plan to work toward inclusion of the PCCC in the Uniform Data System (UDS) and with the OPA/Title X, resulting in public reporting on the measure within six years.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

All facilities that have contributed data to reliability and validity testing have received feedback based on results.

- All partner organizations involved in our nationwide PCCC pilot received reports from the UCSF team containing their facility-level PCCC results and assistance with interpretation. Following advice from the UCSF Institutional Review Board related to provider participant confidentiality in the context of a research study, we did not share provider-level scores with partners. In future usage of the PCCC for quality improvement, this concern will not be relevant.
- UCSF held meetings with staff and leadership of all local clinical partners to discuss overall findings from patient response to the PCCC in their clinics.
- OHA-affiliated clinical sites contributing data to this project received reports containing their facility-level PCCC results from OHA.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

- UCSF staff emailed reports to PCCC pilot partner organization leadership containing a one-page report detailing their facility-level PCCC score and how it was calculated and a list of educational resources for quality improvement. See the Appendix for a sample report to pilot sites.
- UCSF staff facilitated meetings with staff and leadership of local clinical partners to share findings from projects in which the PCCC was used and discuss strategies for performance improvement in contraceptive counseling based on these findings.
- Staff of the central OHA organization emailed reports containing facility-level PCCC results to the leadership of those facilities involved in this project, which then shared these results with facility staff.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

As described in Section 3c.1, we collaborated with partner organizations over the course of the nationwide PCCC pilot to gather input and develop best practices for measure implementation. Feedback was gathered in planning meetings, initial site visits preceding implementation, and Fall 2018 interviews with administrators, providers, and staff on implementation progress and challenges. At the time of results reporting, the UCSF team solicited feedback over email and telephone calls with facility leadership to obtain feedback on the reporting process.

Before the pilot began, we also gathered initial feedback from representatives of clinical organizations that were identified as typical future users of the measure and served as stakeholders in our implementation planning process. Providers and administrators from these organizations participated in Modified Delphi Processes designed to obtain consensus on measure acceptability and feasibility from the clinical perspective. This initial feedback helped guide our collaboration with partner organizations over the course of the pilot.

4a2.2.2. Summarize the feedback obtained from those being measured.

Ongoing communication with partner organizations in planning meetings, site visits, and implementation interviews helped guide implementation planning as described in 3c.1. As described in that section, feedback on measure implementation led to adjustment in the implementation to same-day identification and survey distribution to patients. In our interviews conducted in September 2018, partners described the implementation process as being feasible to integrate into their workflows. When presented with final score reports, leadership of partner organizations expressed overall satisfaction with the utility of results. They expressed that the PCCC score served as a useful insight into the experiences of their patients and would help inform decisions regarding quality improvement efforts. Some partners expressed a desire to be able to compare their PCCC results with those of other organizations. We shared that with future increased use of the measure, measure users would be better able to contextualize results as compared with those of similar organizations using the measure.

4a2.2.3. Summarize the feedback obtained from other users

Participants of Modified Delphi Processes (n=14 administrators, 19 providers) expressed their perceived utility of the measure and its results. Some participants provided feedback on the low feasibility of using EHR to identify patients to respond to the PCCC. This initial feedback was echoed by partner organizations planning to use the measure in the pilot, leading to our development of best practices for in-facility identification of patients and same-day survey administration.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

With regard to the measure itself, we did not receive any feedback that called for any changes to measure specifications. Partner feedback greatly influenced our understanding of best practices for PCCC implementation, as described in Section 3c.1. In addition to helping us develop implementation guidelines, our interactions with partner organizations also helped us develop ongoing feedback processes to encourage consistent data collection and remind staff and providers of data collections goals. Based on partner organization requests for updates on project progress, we instituted monthly feedback reports designed for leadership to share in staff meetings, showing how many surveys had been collected at the clinical site in comparison with data collection goals. In interviews with administrators, providers, and staff about implementation in Fall 2018, we learned that it would be further beneficial to directly share these reports with key staff involved in the project, such as front desk personnel who handed out our surveys, to ensure that they were aware of progress toward project goals.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Because of its limited use during development and testing, we have not yet had the opportunity to collect data demonstrating quality improvement. As described in section 1b.1, the PCCC measure can be used for quality improvement either on its own, or at the same time as the currently endorsed NQF contraceptive measures (#2903 and 2904) to inform quality improvement. In both cases, measurement and reporting of the PCCC measure can be used to identify facilities and/or providers with low performance with respect to patient-centered contraceptive counseling. This information can then be used to target quality improvement activities towards those facilities or individuals in most need of improvement. In the case of facilities, this could involve staff-wide contraceptive

counseling training or the implementation of contraceptive counseling materials (such as digital tools) to facilitate contraceptive counseling. On the individual provider level, this could consist of individual training and capacity building, and/or observation and feedback. Following implementation of these interventions, repeat measurement of the PCCC for the relevant entities can be used to track impact. When used in combination with the currently endorsed NQF measures, the PCCC measure can also serve as a tool for identifying whether attention to method provision is leading to a decrease in attention to patient-centeredness, and allow for intervention if such an effect is observed.

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

An unexpected positive finding during the PCCC pilot was how organizations could use the PCCC survey as a tool to strengthen clinical workflows, helping to improve processes beyond the scope of this project. In one partner clinical facility, there was not a consistent patient check-out process following visits when PCCC implementation was planned to begin, and leadership and staff expressed a desire to institute one. In implementation planning, UCSF staff and partner facility staff planned for patients eligible for the survey to receive paper instructions from medical assistants during their visit, letting patients know to stop at the front desk after their visit to receive a survey. Partner staff used this opportunity to begin using paper instructions with all patients, reminding them to check out after their visit. Patients not eligible for the survey received a different color piece of paper than those who were eligible, allowing front desk personnel to distinguish eligibility. This process was successfully implemented with all patients, and thus a tool designed to support PCCC data collection helped the partner organization accomplish other administrative goals. Both partner organizations and patients benefited, with patients having the additional opportunity to understand any billing needs and address any other administrative concerns with the facility during check-out.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

In addition to unexpected positive benefits to clinic workflows, we also learned that patients may have directly benefited from the opportunity to respond to a measure about their patient experience of contraceptive counseling. Some staff tasked with distributing the survey to patients (e.g., front desk personnel) reported in implementation interviews conducted in Fall 2018 that some patients expressed unprompted verbal appreciation to them for the opportunity to take the survey, and that they liked the questions on the survey. Therefore, one unexpected benefit of implementation may be an increased sense of agency and engagement among patients.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):

Are the measure specifications harmonized to the extent possible?

No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

N/A

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

There are no other measures assessing the same specific area of focus or target population (patients who received contraceptive counseling). However, we wish to acknowledge two measures with conceptual overlap to the PCCC: CG-CAHPS (NQF measure #0005) and the OPA-developed measures for contraceptive provision (NQF measures #2903 and 2904).

Both the PCCC and CG-CAHPS are PRO-PMs concerned with patient experience and particularly provider-patient communication. While there are similarities between how the PCCC and the CG-CAHPS communication subscale conceptualize this communication, CG-CAHPS is a general measure applicable to many care contexts and the PCCC is designed specifically for the unique context of contraceptive counseling. The choice of a contraceptive method is a highly preference-sensitive decision with many possible outcomes – most patients choose between more than ten methods that are medically appropriate for them. Each patient has their own preferences for what is most important to them in a contraceptive method (e.g. pregnancy prevention, minimal side effects, control of menstrual bleeding), and what is preferable with regard to those priorities (e.g. having a monthly period or having no period). Thus, each individual has their own unique preference profile, and patient-centered contraceptive counseling as measured by the PCCC is focused on these individualized preferences and attentive to the highly personal and sensitive nature of discussion and decision making around sex and pregnancy. The PCCC is purposely designed with input from patient and provider stakeholders to address this specific context of the contraceptive counseling conversation. The PCCC's focus on the domains of adequate contraceptive information, decision support for a complex, preference-sensitive decision, and interpersonal connection on this personal topic distinguishes the PCCC from CG-CAHPS. The distinction between the two measures was echoed in our communications with patients about this topic. During the course of our process of developing and validating our PCCC measure, we explored with our patient stakeholder group their feelings about the relationship between the CG-CAPHS measure and PCCC. They confirmed the importance of a measure specific to contraceptive care for the reasons outlined above.

While unrelated, the contraceptive provision measures are the only other NQF-endorsed measures to address quality in the context of family planning care. As described in Section 1b.1, an original motivation for PCCC development was the need for a PRO-PM of patient-centered contraceptive counseling to counter-balance use of

the contraceptive provision measures. When used together, these measures can provide a robust picture of contraceptive care quality, and ensure that advances in contraceptive provision do not come at the cost of patient experience.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** NQF_Appendix_UCSF_PCCC_2Oct2019.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): University of California, San Francisco

Co.2 Point of Contact: Christine, Dehlendorf, christine.dehlendorf@ucsf.edu, 628-206-8712-

Co.3 Measure Developer if different from Measure Steward: University of California, San Francisco

Co.4 Point of Contact: Christine, Dehlendorf, christine.dehlendorf@ucsf.edu, 628-206-8712-

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Christine Dehlendorf, University of California San Francisco. Dr. Dehlendorf is a family physician and reproductive health researcher. She initially developed the Patient-Reported Outcome Measure (Interpersonal Quality of Family Planning scale, or IQFP), adapted it to a shorter four-item version, and then led the validation of this measure as a performance measure.

Danielle Hessler Jones, University of California San Francisco. Dr. Hessler is a psychologist and psychometrics expert. She acted as a co-Investigator for the development of the IQFP as well as its adaptation and testing of the validity and reliability of the performance measure.

Eric Vittinghoff, University of California San Francisco. Dr. Vittinghoff is an expert biostatistician with a broad range of analytic expertise. He acted as co-investigator and led the statistical analysis for validity and reliability testing.

Kelsey Holt, University of California San Francisco. Dr. Holt, a public health expert with expertise in contraceptive care, acted as a consultant for the team and lent her expertise in developing a patient-reported measure of contraceptive counseling.

R. Adams Dudley, Philip R. Lee Institute of Health Policy Studies. Dr. Dudley, a physician, with experience with performance measure development and validation, acted as co-investigator for the development of the measure and lent his expertise in patient-reported measure development.

Elizabeth Jones, National Family Planning and Reproductive Health Association (NFPRHA). Ms. Jones, Director of Service Delivery Improvement at NFPHRA, assisted with identification of sites to test the measure and conducted interviews with clinic administrators for feedback around implementation.

Daryn Eikner, National Family Planning and Reproductive Health Association (NFPRHA). Ms. Eikner, Vice President of Service Delivery Improvement at NFPRHA, acted as an organizational stakeholder to ensure that provider and clinic perspectives were represented in the development of the measure.

Edith Fox, University of California San Francisco. Ms. Fox acted as the Project Manager for the development of the measure and oversaw data collection activities and coordination with participating clinics.

Reiley Reed, University of California San Francisco. Ms. Reed acted as a Project Coordinator for the development of the measure, coordinated with clinic sites during data collection and conducted interviews with clinic administrators, providers, and staff around implementation of the measure.

Ilana Silverstein, University of California San Francisco. Ms. Silverstein acted as a Project Coordinator and managed incoming data from participating sites during the testing phase.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released:

Ad.3 Month and Year of most recent revision:

Ad.4 What is your frequency for review/update of this measure?

Ad.5 When is the next scheduled review/update for this measure?

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: Additional UCSF steward/developer contact: Ilana Silverstein, Project Manager. ilana.silverstein@ucsf.edu. Phone: (628)206-5092.